Structural and pharmacological study of Trypanosoma cruzi trans-sialidase

Chen, Zexin

Award date:
2017

Awarding institution:
University of Bath

Structural and pharmacological study of *Trypanosoma cruzi* trans-sialidase

Zexin Chen

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

October 2017

Signed.....

Date.....

Acknowledgements

Firstly, I would like to express biggest appreciation to my lead supervisor, Dr. Susan Crennell, for the meticulous guidance and encouragements of my researches, especially in structural biology and bioinformatics studies through these long years, also, great help in proof reading my thesis. I would like to thank my second supervisor, Prof. Jean van den Elsen for the great ideas and supports in biochemistry and biophysics. I would like to say thank you to my third supervisor, Dr. Andrew G. Watts, who devoted a large amount of energy for my project, and provide concentrated supports in many areas, especially in chemistry.

I would like to thank Prof. Ravi Acharya and his group members, for the help in X-ray crystallography, including sample preparation, data collection and processing. I would like to thank Prof. Stephen Husbands for numbers of helpful advice in chemistry and warm encouragements. I would like to acknowledge all the input that Dr. Tim Woodman gave with my NMR spectra, and Dr. Anneka Lubben for her contribution and technical help regarding mass spectrometry instrumentation and data processing. I would like to thank Dr. Stefan Bagby, for the help in cloning and molecular biology.

I would like to thank for the help from all the lab members in lab 4S 0.27 and 5W 3.21 (Yi, Mareike, Abhishek, Ayla, Chaterine, Marjorie, Ricardo, Terrence, Andreas and Bayan) and all the project students throughout these years.

I would like to thank Diamond Light Source for access to beamline I02, I03 and I04 that contributed to the results presented in this thesis.

I would like to say thank you to the University of Bath and Great Britain-China Educational Trust for the financial support.

I would like to give the biggest thank you to my family for all the love, support,

encouragement that they gave during the last 28 years!

Last but not least, I'd like to not name individual people, instead will just leave a big thank to all of my friends that, wherever they are, gave me the strength during PhD, especially to survive the period of writing up. Thank you!

Abstract

Chagas disease affects millions of people in Latin America and is now spreading to the whole world. Causing by the infection of *Trypanosoma cruzi* (*T. cruzi*), Chagas disease can be fatal, patients suffering from severe intestinal and cardiac disorders. The development of diagnosis and therapeutic drugs against Chagas disease is slow so far due to limited investment. *T. cruzi* trans sialidase (TcTS) is a surface-anchored glycoside transferase which plays a critical role in host cell invasion and immune evasion. Its uniqueness and role in parasite survival and pathogenicity indicates the importance of TcTS as a potential target for therapeutic drug development. However, no sufficiently strong inhibitors of TcTS have been described. Based on the structure of sialyl-lactose, a natural substrate of TcTS, we have designed a new type of inhibitor, α-aminophosphonate, which inhibits TcTS effectively and selectively when compared to inhibition of bacterial sialidases/neuraminidases.

Interestingly, although the compound was designed to interact with the catalytic site, it showed noncompetitive inhibition. Crystal structures suggest an allosteric binding site and the interaction between TcTS and α-aminophosphonate causes only minor conformational changes. Binding of α-aminophosphonate may inhibit TcTS activity by inducing these conformational changes and increasing the structural stability of catalytic site This results in stable binding of lactose, which is also shown through thermodynamics. Additionally, anthraquinone derivatives, whether an individual compound or as a part of an α-aminophosphonate, inhibit TcTS non-competitively, but these are shown to bind to the lactose-binding site, part of the catalytic pocket, also giving new insight into the inhibition and catalytic mechanisms of TcTS.

In conclusion, α-aminophosphonate may be a potential drug candidate for Chagas disease. The crystal structure of TcTS in complex with α-aminophosphonate compound is the first report of the TcTS allosteric binding site, which may be crucial for better

understanding of the mechanisms and substrate recognition of this vital enzyme and for rational drug design.

Table of contents

List of Figures

List of Tables

List of abbreviations

Abbreviation	*Meaning*
Abs	Absorbance
AcOH	Acetic acid
Arg	Arginine
Asp	Aspartic acid
br	Broad
Brine	Saturated solution of sodium chloride
CD_3OD	Tetradeuteromethanol
$CDCl_3$	Deuterated chloroform
Conc.	Concentration
d	Doublet
DCM/ CH_2Cl_2	Dichloromethane
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
eq.	equivalent
EtoAc	Ethyl acetate
F.T.	Flow through
Glu	Glutamine
GPI	glycosylphosphatidylinositol
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HMBC	Heteronuclear multiple bond coherence
hr	Hours

Abbreviation	*Meaning*
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single quantum coherence
Hz	Hertz
IC_{50}	Half maximal (50%) inhibitory concentration (IC)
Ile	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
J	Coupling constant
k_{cat}	Turnover number
K_d	Binding constant
K_i	Inhibition constant
K_m	Michaelis constant
Leu	Leucine
m	Multiplet
M	Molar
MeOH	Methanol
MHz	Megahertz
min	Minutes
mM	Millimolar
Mw	Microwave
NMR	Nuclear magnetic resonance
o/n	Overnight
occ.	Occupancy
ºC	Degree Celsius

Abbreviation	*Meaning*
PEG	Polyethylene glycol
ppm	Parts per million
q	Quartet
r.t.	Room temperature
s	Singlet
SA	Sialic acid
SAL	Sialyllactose
SAR	Structure-activity relationship
sec	seconds
Ser	Serine
t	Triplet
T. brucei	*Trypanosoma brucei*
T. cruzi	*Trypanosoma cruzi*
TcTS	*T. cruzi* trans-sialidase
Thr	Threonine
Tris	Tris(hydroxymethyl)aminomethane
Trp	Tryptophan
TS	Trans-sialidase
Tyr	Tyrosine
Val	Valine
wt	Wild type
XRC	X-ray crystallography
δ	Chemical shift

Chapter 1. Introduction

1.1. *Trypanosoma cruzi* and Chagas disease

1.1.1. Trypanosomes and Trypanosomiasis

Trypanosoma, the name coming from the Greek trypano (borer) and soma (body), is a genus of unicellular protozoa belonging to the class of kinetoplastids (Atwood *et al*., 2005). All *Trypanosoma* are parasitic and can be found in insects, some species can infect mammals and cause disease, including the fatal human disease. African sleeping sickness, which is caused by *Trypanosoma brucei* (*T. brucei*), which normally transferred by the tsetse fly, and Chagas disease that is induced by *Trypanosoma cruzi* (*T. cruzi*) which is transmitted by triatominae (Figure 1.1) (Podlipaev, 2001). *T. rangeli* can also infect humans via triatominae but it shows no evidence of human pathogenesis (Stoco *et al.,* 2014). Other types of trypanosomes primarily infect mammals other than humans, for example, *T. congolense*, *T. simiae* and *T. vivax* cause nagana, also known as animal trypanosomiasis, in different kinds of mammals (Wells, 1972).

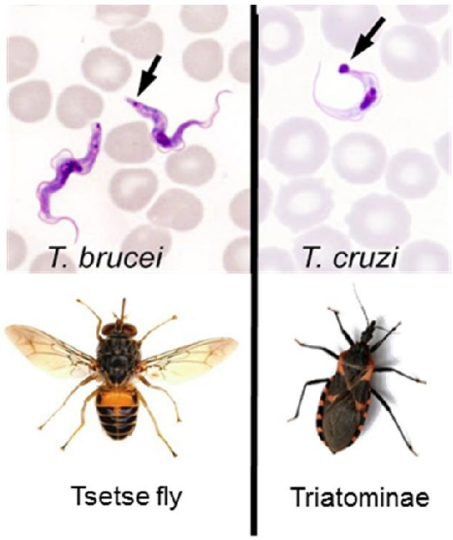


Figure 1.1. Two major trypanosomes causing threatening human health: *T. brucei* (left) and *T. cruzi* (right), together with their carriers: tsetse fly and triatominae, respectively (adapted from Pritt, 2011).

African sleeping sickness and Chagas disease are the two types of trypanosomiasis which impact humans. African sleeping sickness threatens 36 sub-Saharan countries

with about 10,000 reported cases in 2009, but large numbers of undiagnosed cases occur and about 2000 new infections have been reported each year (WHO, 2017). Chagas disease is more lethal and caused more than 10,000 deaths annually (Stanaway and Roth, 2015)

1.1.2. Chagas disease — a global health challenge

Chagas disease, also known as American trypanosomiasis, was first discovered and named by the Brazilian bacteriologist Carlos Chagas in 1909 (Chagas, 1909). During the 20th century, huge efforts towards understanding the disease have been carried out including identification of *T. cruzi* and its life cycles, investigation of *T. cruzi* transmission and treatments of Chagas disease (Steverding, 2014a). However, over a century following its identification Chagas disease still threatens public health all around the world and many details of the pathogenesis, for example, the relationship between *T. cruzi* infection and organ failures in the chronic phase of Chagas disease, are still unclear. Furthermore, although several anti-protozoan drugs have been developed for the treatment of Chagas disease, they still only have limited therapeutic effect on the chronic stage of Chagas disease, and poor early diagnosis is also an issue.

Chagas disease is mainly endemic in Latin America, but today, it has spread all around the world, cases of *T. cruzi* infected patients have been reported in North America, Europe, Australia and Japan (Figure 1.2) (Coura and Vinas, 2010). The total number of reported cases of Chagas disease all around the world is about 7-8 million (Pereira and Navarro, 2013). In endemic regions, *T. cruzi* is primarily transmitted by triatominae, while over 100 species of triatominae have been shown to transmit *T. cruzi*, 4 species are known to be the most important in epidemiology, including *Triatoma infestans* which is the most intensively studied (Cecere *et al.*, 1999). Mammals, especially domestic mammals such as dog, cat, pig and goat, can also be infected by *T. cruzi* (Noireau *et al*., 2009). Whilst being victims of *T. cruzi* infection, these mammals also play important roles in the transmission of Chagas disease. Oral transmission was also

reported where patients were infected by *T. cruzi* through eating contaminated fruit (Nóbrega, 2009). The spreading of Chagas disease beyond Latin America is mainly attributed to migration of population from the endemic region to the whole world. In the non-endemic region, Chagas disease can be transmitted through blood transfusions, organ transplantation, mother-to-child transmission, and accidental laboratory exposure (Guhl *et al.*, 2000).

In endemic countries, mostly developing countries, Chagas disease impacts public health and causes substantial financial burdens. The global cost of Chagas disease is US$7.2 billion annually, which exceeds the expenditure on many other severe diseases (compared with US$ 2 billion on rotavirus and US$ 4.7 billion on oral cancer) (Lee *et al.*, 2013).

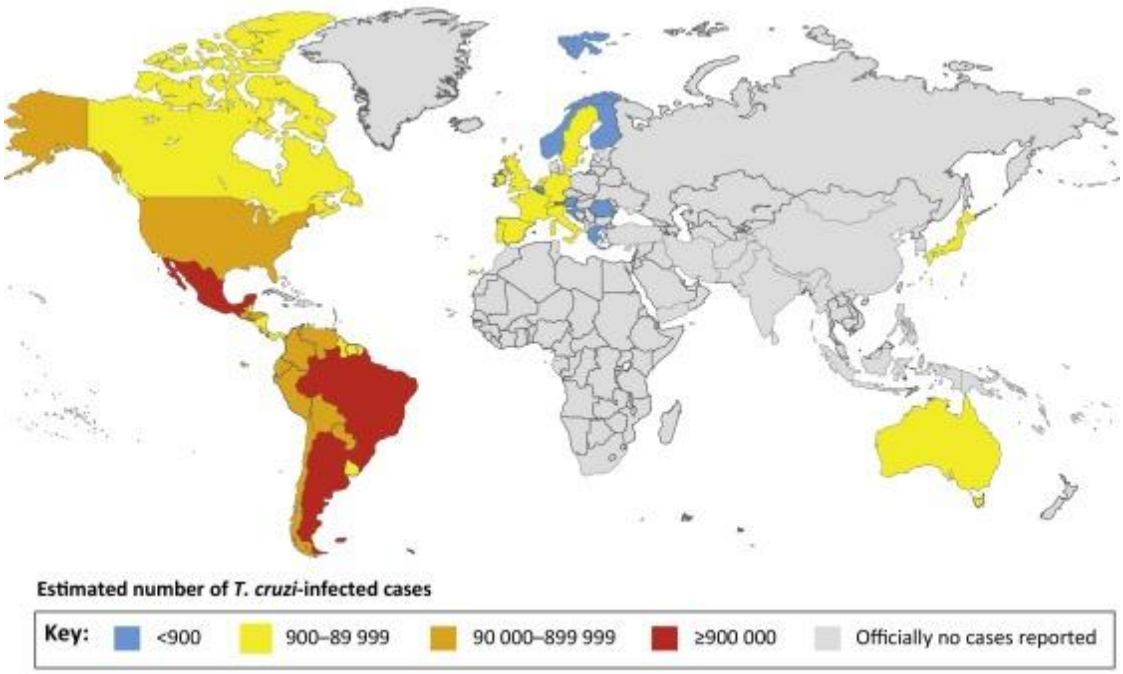


Figure1.2. Distribution of Chagas disease (Perez *et al.*, 2015)

The life cycle of *T. cruzi* has been comprehensively reviewed by Tyler and Engman (Tyler and Engman, 2001). During its life cycle, *T. cruzi* moves between insect hosts (triatominae) and mammal hosts include human (Figure 1.3). In insect hosts, the parasites multiply and differentiate to metacyclic trypomastigotes (MT), which is an infective stage (CDC, 2014). The MT then can enter human peripheral blood via a sting by triatominae, are transported to different tissues and organs and invades the host cells. After amplification in host cells, the trypomastigotes differentiate into amastigotes, which can develop to trypomastigotes and enter the bloodstream through lysis of the host cell.

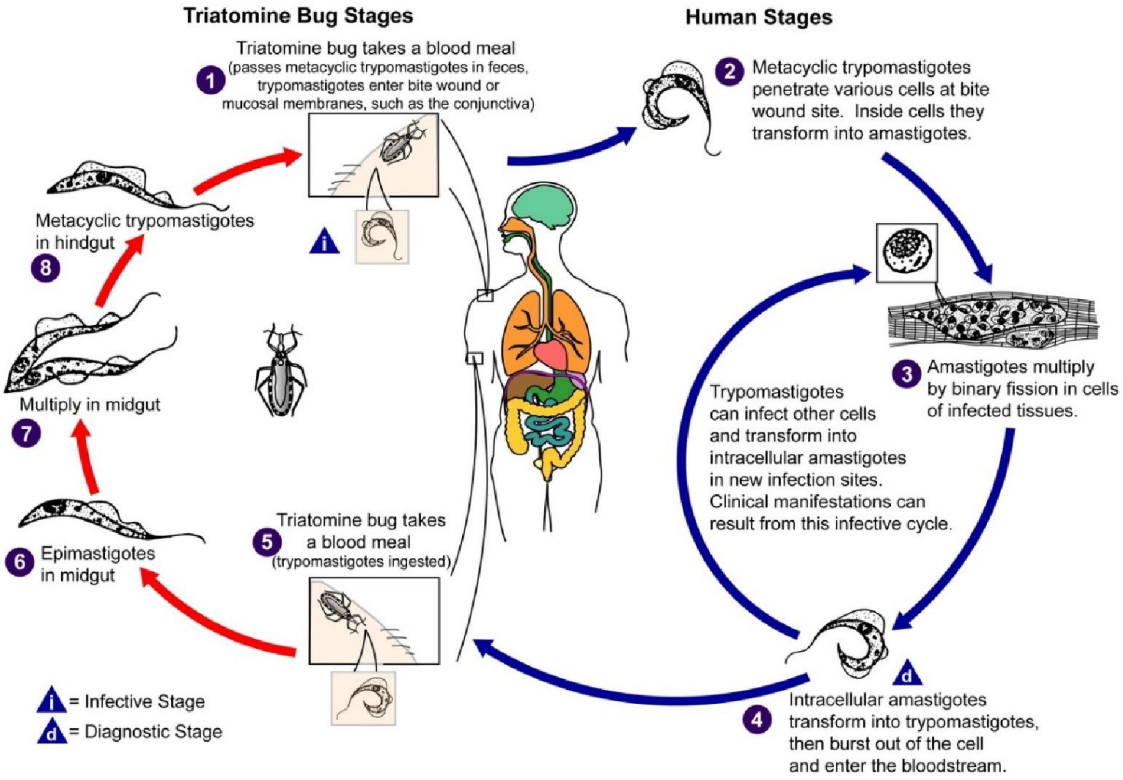


Figure 1.3. Life cycle of *T. cruzi* (Clayton, 2010)

The clinical aspects of Chagas disease can be divided into two stages: the acute and chronic phases. Regarded as a "silent killer" (Maguire, 2006), many patients that catch Chagas disease present mild symptoms during most of the time of pathogenesis processes. In the acute stage of Chagas disease, over 50% of patients could be asymptomatic, and the symptoms presented in acute phase tend to be non-specific, such

as fever, subcutaneous edema, and enlargement in liver, most of these symptoms will resolve in several months spontaneously (Rassi *et al.,* 2010). Some inflammatory manifestations are also used to identify acute Chagas disease, for example, unilateral periorbital edema, also called Romaña sign (Figure 1.4 A), which is caused by conjunctival infection through rubbing eyelids (Rassi *et al.*, 2010). Even with treatment against infect parasites, over 90% of patients will progress to the chronic phase during the development of the disease (Bern *et al.*, 2007), which is more concealed and lethal than acute phase Chagas disease. 30% of chronic patients suffer from digestive problem such as megaesophagus and secondary achalasia (Rassi *et al.*, 2010; Salamanca-Dejour *et al*., 2012). Around two thirds of chronic Chagas disease patients will develop cardiac disorders including heart damage (Figure 1.4 B), and Chagas disease may induce sudden death through by heart rate abnormalities (Machado *et al.*, 2012). Another manifestation is neurological symptoms, e.g. neuritis and dementia, caused by neuronal invasion by *T. cruzi* (Cordova *et al*., 2010).

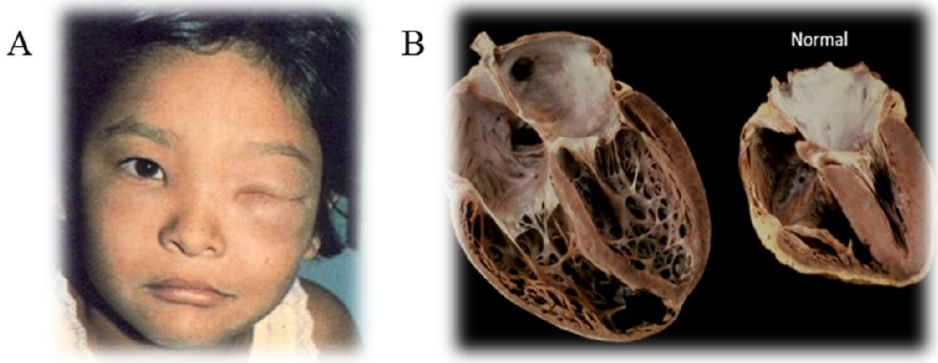


Figure 1.4. (A) Romana's sign and **(B)** Heart of Chagas patient compared with normal heart. (accessed from http://pharmeasy.in)

Since the discovery of Chagas disease, many efforts have been made to find effective chemical prevention and therapeutic agents. Lindane and Deltamethrin were found to kill triatominae bugs thus they can prevent the transfer of Chagas disease (Steverding, 2014). The first compound that was identified that killed *T. cruzi* was crystal violet, which is normally used as a dye for bacteria classification. Later in the 1960s to 1970s,

Benznidazole and Nifurtimox were discovered as two effective anti-parasite drug against *T. cruzi* and these are the only drugs available for Chagas disease medication today (Steverding, 2014). However, both are still controversial due to their severe adverse side effects (Castro *et al.*, 2006). Benznidazole can lead to hypersensitivity, polyneuropathy and bone marrow problems (Pinazo *et al.*, 2010). Nifurtimox can also induce anorexia, psychic alterations, excitability, and sleepiness (Castro *et al.*, 2006). So far, there is also no commercially available vaccine against Chagas disease (*Lee et al.*, 2010). More importantly, both Benznidazole and Nifurtimox have better effect in early diagnosed cases and only show limited effect on chronic Chagas disease (Hamers *et al*., 2016; Morillo *et al.*, 2015; Munoz *et al.,* 2013). Therefore, there is an urgent need to develop an effective, specific and safe drug against Chagas disease.

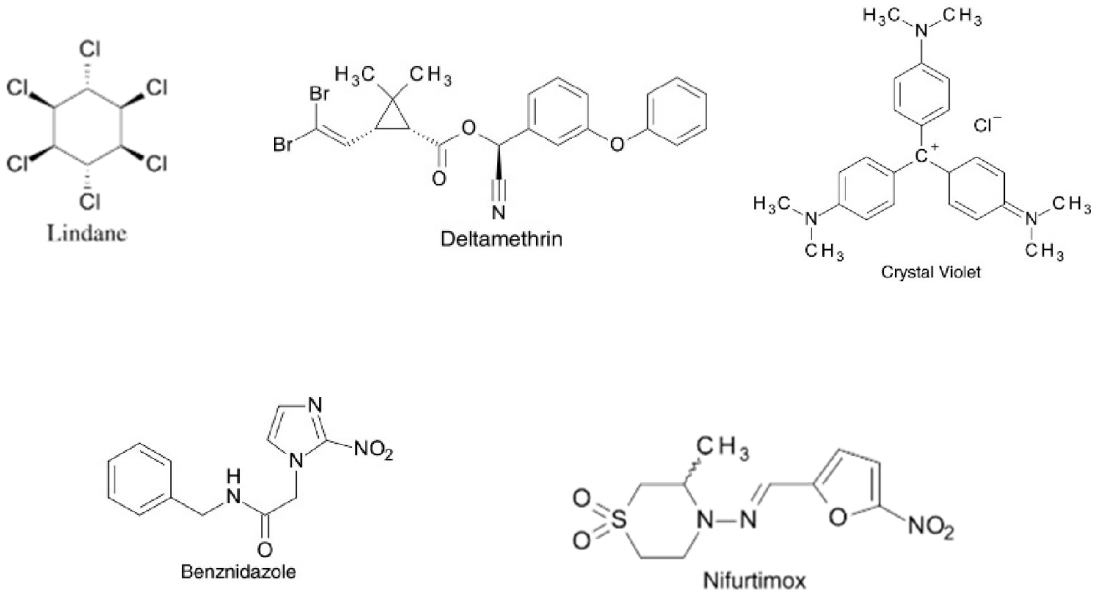
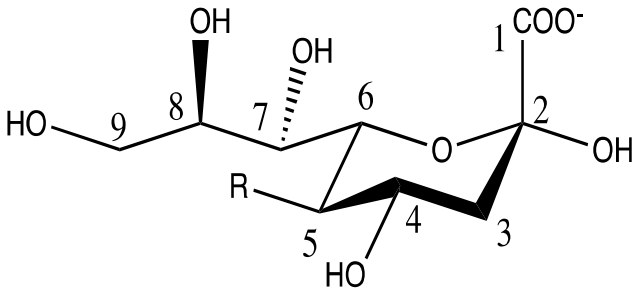


Figure 1.5. Structures of current therapeutic drugs against Chagas disease

1.2. Sialic acid and *T. cruzi* trans sialidase

1.2.1. Sialic acids

In the broad sense, the term sialic acid (SA), also known as neuraminic acid, covers a family of N- or O-substituted 9 carbon acidic sugar derivatives which can be a part of glycans that can be found in both eukaryotic and prokaryotic cells. More precisely, SA is often regarded as the name of the most common member of this family, 2-keto-3-deoxy-5- acetamido-D-glycero-D-galacto-nonulosonic acid, simply called N-acetylneuraminic acid (Neu5Ac or NANA) (Figure 1.6.) (Varki, 2009), Throughout this thesis SA will represent Neu5Ac unless defined otherwise. In some cases, two other members of this family, 2-keto-3-deoxynononic acid (KDN) and N-glycolylneuraminic acid (Neu5Gc) are also called SA (Figure 1.6.). The substituents on C4, C5, C7, C8 and C9 positions lead to more than 50 members of the wider SA family (Varki, 2009).



Neu5Ac: R= CH_3-CO-NH-

KDN: R=OH

Neu5Gc: R= $HOCH_2$-CO-NH-

Figure 1.6. Structure of sialic acid

Neu5Ac is the most ubiquitous member of the SA family in nature. After first isolation from bovine submaxillary gland in 1936, the structure, synthesis and function of Neu5Ac has been studied intensively in the last century. Neu5Gc differs from Neu5Ac only by a single hydroxyl group at the end of the *N*-acetamido group. Neu5Gc is the second ubiquitous SA in nature following Neu5Ac, however, it is not present in humans

(Varki, 2010). Therefore, the human immune system normally recognizes Neu5Gc as a foreign molecule. It is also reported that binding between Neu5Gc and its antibody might induce chronic inflammation and be correlated to human cancer (Coussens and Werb, 2002; Samraj *et al*., 2014; Tan and Coussens, 2007). KDN is a relatively newly discovered molecule in the SA family having been studied for around 20 years. KDN mainly occurs in bacterial cells with the similar occurrence level of Neu5Ac, however in human body it is only enriched in fetal erythrocytes and tumor cells rather than healthy cells (Inoue and Kitajima, 2006).

As negatively charged carbohydrates, SAs are hydrophilic. Sialic acid can decorate almost all kinds of glycans as the terminal sugar, by formation of glycosidic bond via C2 (3'-sialylglycan) or C4 (6'-sialylglycan), various substituents can also link at the C4, C5, C7, C8 and C9 position (Figure 1.7) (Angata and Varki, 2002; Arnaud *et al*., 2013; Varki and Varki, 2007). Apart from the presentation and structure, the universality and diversity of SA on both human and microbial cell is shown from its many functions. An essential function for SA is to maintain membrane stability and modulate interactions between cell and environment, SA also paly important role in protecting cell surface antigens and regulate the immune response (Wiederschain, 2009). The essential functions of SA are often exploited by cancer cells and infecting microbes. By significant overexpression of SA-rich glycoproteins on the cell membrane, cancer cell can generate a strongly negative-charged surface which can lead to cell opposition and enhance cancer cell metastasis (Fuster and Esko, 2005).

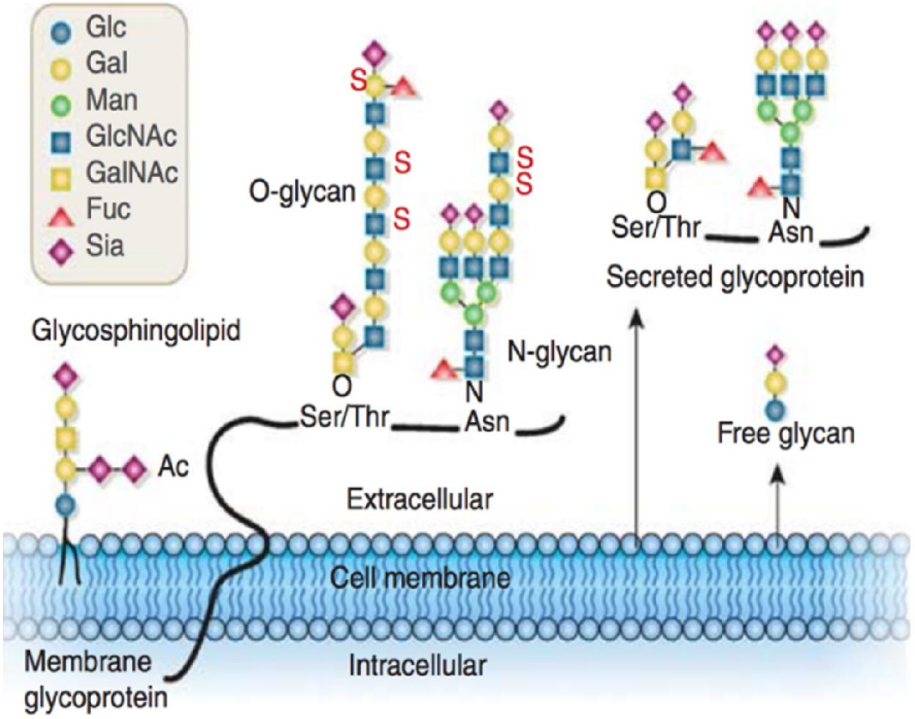


Figure 1.7. Diversity of sialic acid decorated glycans on cell surface (Varki and Varki, 2007)

SA also has critical roles in infection and pathogenicity of infecting microbes including bacteria, parasites and viruses. Many bacteria and parasites can incorporate SA on their surface for evasion from immune surveillance (Hicks *et al.*, 2000; Severi *et al.*, 2007). In virus infection, SA contributes to cell attachment and regulation of membrane penetration (Stencel-Baerenwald *et al.,* 2014).

1.2.2. *T. cruzi* trans sialidase

1.2.2.1. Structure and function

T. cruzi trans-sialidase (TcTS) is a critical enzyme located on the surface of the parasite. The SA terminated mucin was discovered to play an important role in host cell invasion of *T. cruzi* in 1991 (Schenkman *et al.*, 1991), from where the SA hydrolyzed and transferred to the surface of the parasite. Later, the SA transfer and hydrolysis activity were proved to be mediated by the same enzyme, which has been identified as TcTS

(Schenkman *et al*., 1992). TcTS belongs to the glycoside hydrolase GH33 family, which contain most of prokaryotic and some eukaryotic sialidases/neuraminidases (Amaya, *et al.*, 2004; Henrissat and Davies, 1997). TcTS can be divided into two different parts (Figure1.8 A), the N-terminal functional part and the C-terminal shed acute-phase antigen (SAPA) domain.

The SAPA domain is a non-structural repeat of a 12-residue unit (DSSAHGTPSTPV), which is not involved in the catalytic activity, but can increase immune response towards TcTS (Buscaglia *et al*., 1999; Buscaglia *et al*., 1998). The C-terminus of the SAPA domain links TcTS to the cell surface of *T. cruzi* by a Glycosyl-phosphatidylinositol (GPI) anchor (Agusti *et al*., 1997). The N-terminal part of TcTS contains a 6-stranded β-propeller catalytic domain, which is structurally similar to other members of GH33 family, and a β-sandwich lectin-like domain (Figure 1.8B) (Buschiazzo *et al.,* 2002). The catalytic domain directly contributes to the trans-glycosylation activity from host glycoconjugate to parasite surface mucin, through the catalytic site located in the centre of the β-propeller (Figure 1.8 B). A lectin-like domain has been found in several sialidases, for example, the sialidases from *Vibrio cholerae, Streptococcus pneumoniae*, and trans-sialidase from *Trypanosoma congolense* (Crennell *et al*., 1994; Waespy, Mario *et al*., 2015). In most cases, the sialidase lectin-domain shows sugar binding activity and has been proved to be involved in substrate recognition and catalytic activity (Moustafa *et al*., 2004; Waespy, M. *et al*., 2015; Yang *et al*., 2015). However, there has been no evidence to show that the TcTS lectin-like domain has a role in catalytic activity, or sugar binding activity.

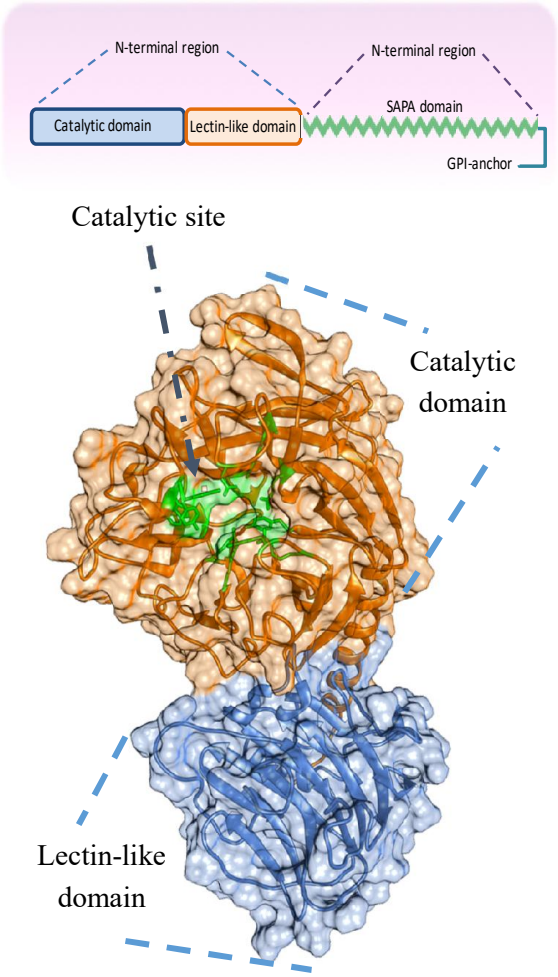


Figure 1.8. **A.** Domains of TcTS; **B.** Crystal structure of TcTS, the catalytic domain is presented in orange and the catalytic site is in green, the lectin-like domain is presented in blue (PDB. 1MS5) (adapted from Buschiazzo *et al.,* 2002).

T. cruzi cannot synthesize SA, therefore, the transfer activity of TcTS is the only source of sialic acid for *T. cruzi*. Additionally, an *in vivo* study suggests that injection of TcTS in infected mice enhances parasitemia symptoms and mortality rate (Freire-De-Lima *et al*., 2012). Therefore, TcTS is critical for the pathogenicity and survival of *T. cruzi*. Using TcTS activity, *T. cruzi* can decorate the terminus of its mucin glycans with SA, in order to bury the surface antigen, thus evade attack by anti-galactosyl antibodies and the complement system (Morrot, 2013).

Moreover, SA is believed to play important roles in TcTS cell invasion. A recent

hypothesis for the molecular mechanisms of TcTS in the cell invasion process is that sialylation of the parasite surface glycans by TcTS produces ligands for SA binding Ig-like lectins (siglecs) on host cells (Jacobs *et al*., 2010). By interacting with sialylated glycans, siglecs can increase cell-cell interaction, and adherence of the parasite to the host cell (Varki and Gagneux, 2012). Activation of siglecs also modulates the innate immune response which would be beneficial to the persistence of *T. cruzi* in the host (Freire-de-Lima *et al*., 2016). However, Yoshida's group reported that removal of the SA from *T. cruzi* surface resulted in even stronger cell-cell interactions, arguing that SA is not necessary for cell invasion (Yoshida *et al*., 1997). However, inhibition or decrease of TcTS levels on the *T. cruzi* surface was reported to eliminate trypomastigote virulence (Agusti *et al*., 2004; Freire-De-Lima *et al*., 2012; Jacobs *et al*., 2010; San Francisco *et al*., 2017), which suggests that some undetermined mechanism, not the catalytic process, is responsible for the contribution of TcTS to cell invasion by trypomastigotes. In the interaction between parasite and the host cell, TcTS hydrolyzes host cell surface glyconjugates, result in exposure of the β-Gal*p* residue, which could interact with TcTS and TcTS-like glycoproteins. Meanwhile, TcTS and TcTS-like glycoproteins can also bind to the sialyl-glycans on parasite surface, increasing the adherence between *T. cruzi* and host cell (Varki and Gagneux, 2012). The absence of the SA residue and exposure of β-Gal*p* also increases penetration of host cell surface, which creates favorable conditions for parasite invasion (Monteiro *et al.*, 1998). By interacting with the immune cells, TcTS can favour the immune evasion of *T. cruzi* whilst, it can also contribute to infection of the parasite by cellular interactions.

TcTS and its trans-sialylation activity can also induce other host cellular processes apart from cell invasion. As a critical factor in *T. cruzi* immune evasion, TcTS contributes to the immune response by interaction with a number of different immune cells. The SAPA region of TcTS has been shown to cause T-cell independent B-cell activation and production of non-specific antibodies, which protects *T. cruzi* from specific attack (Gao *et al.,* 2002). Additionally, both active and inactive TcTS can also activate $CD4^+$ T cells, and block apoptosis, suggesting that this process is unrelated to TcTS catalytic activity

(Todeschini *et al.,* 2002). For cytotoxic $CD8^+$ T cell against *T. cruzi*, TcTS can mask the surface β-Gal*p* residues through its SA transfer activity and thus decrease $CD8^+$ T cell activity (Freire-De-Lima *et al.,* 2010). It is reported that TcTS activates the NF-κB signaling pathway of endothelial cells, which will prevent the cell from undergoing apoptosis and increase susceptibility to parasite infection (Dias *et al.,* 2008). A imilar function of TcTS was also found in neuron and Schwann cells (Chuenkova *et al.*, 2001; Chuenkova and Perrin, 2004). Moreover, TcTS can remove SA residues from platelet surfaces then induce platelet clearance (Tribulatti *et al*., 2005)

As well as being anchored on the *T. cruzi* surface, TcTS can also be secreted by the parasite. It is worth noting that secretion of TcTS is not the release of single protein molecules, but through shed vesicles formed of plasma membrane (Agusti *et al.*, 2000). The shed TcTS can achieve multiple biological functions without direct parasite attachment. Injection of these TcTS-rich shed vesicles into mice induces severe heart damage and inflammatory responses (Torrecilhas *et al.*, 2009), suggesting that shed TcTS participates in multiple pathogenicities of *T. cruzi* infection including the Chagas heart disease. The majority of cellular function studies of TcTS were studied using purified recombinant TcTS due to the lack of a strong inhibitor or knock-out *T. cruzi* strains (Freire-de-Lima *et al.,* 2016). Thus, these functions can be classified into the functions of shed TcTS until the experimental evidence of surface-anchored TcTS is presented.

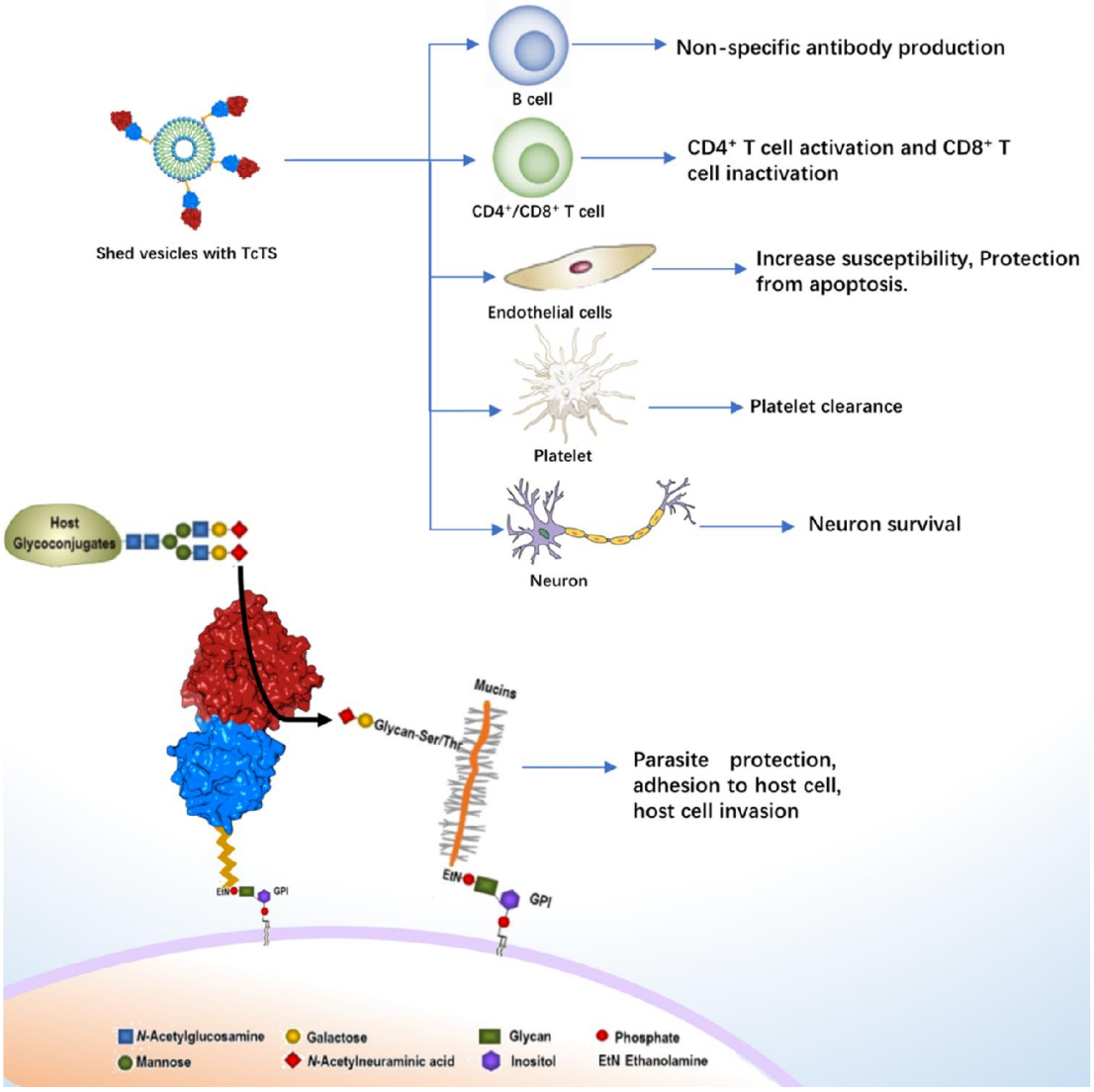


Figure 1.9. Functions of TcTS (adapted from Buscaglia *et al.*, 2006; Nardy *et al.*, 2016)

In conclusion, TcTS is involved in multiple processes during *T. cruzi* infection, it can play a role as SA modifier, immune modulator and even cell adhesion agent. TcTS performs its functions either anchored on the parasite surface or shed into the plasma. However, the lack of an effective and selective inhibitor or a gene modified *T. cruzi* model lead to difficulties in the study of TcTS function. Therefore, in many cases, it is still unclear whether the functions of TcTS are domain-dependent or catalytic activity dependent, and the molecular mechanisms of many functions are still unknown. The development of a new generation of TcTS inhibitors might favor intensive studies of the molecular details of this critical enzyme and further our understanding of Chagas disease pathogenicity and treatments.

1.2.2.2. Substrate and catalytic mechanisms

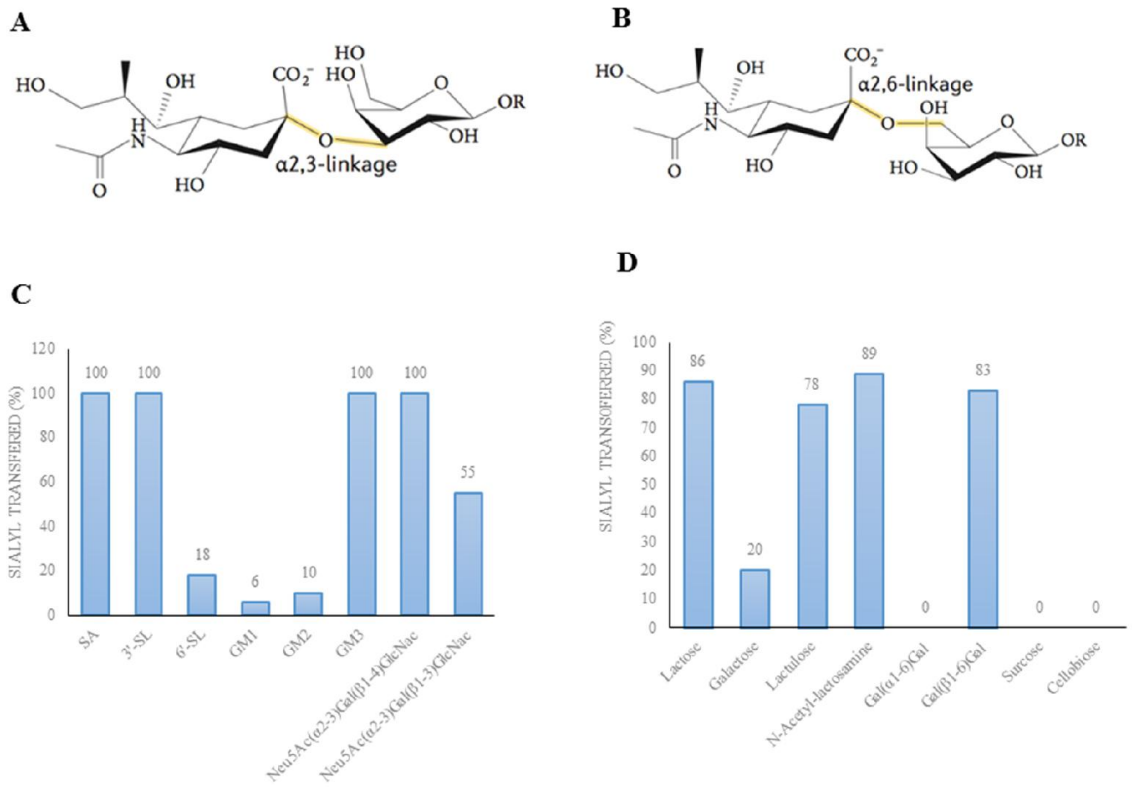


Figure 1.10. A. Structure of Neu5Ac-α2,3-Gal and **B.** Neu5Ac-α2,6-Gal **C.** Rate of sialyl transfer with different donor substrate. **D.** Rate of sialyl transfer with different acceptor substrate (adapted from Vandekerckhove *et al.*, 1992)

TcTS is highly selective in its substrates for both hydrolysis and sialylation. As a donor substrate, TcTS specifically hydrolyzed Neu5Ac-α2,3-Gal (Figure 1.10A) rather than Neu5Ac-α2,6-Gal (Figure 1.10B) (Vandekerckhove *et al*, 1992). However, any Neu5Ac-α2,3-Gal terminal glycan can be the SA donor for TcTS (Figure 1.10 C), for example, ganglioside GM_3. A terminal galactose residue is critical for an acceptor substrate, and only Gal-β- glycans are able to be sialylated by TcTS (Figure 1.10 D), but a single galactose is only a weak acceptor substrate for TcTS (Vandekerckhove *et al.*, 1992)

The interaction between TcTS and sialyl-lactose (SAL) is a two-step displacement

reaction following a "ping-pong" mechanism (Figure 1.11) (Damager *et al*., 2008). Initially, the donor SAL would enter the catalytic site of TcTS, leading to a conformational change of Tyr119 (Figure 1.12 A, B) which is then able to form the stacking interaction of the galactose binding site. The sialyl moiety would also fit into the sialyl binding site including the binding of the carboxyl group with the enzyme's arginine triad and hydrogen bond formation between the 4-OH and Arg53, the acetamido side chain and Asp96, and the glycerol side chain and Trp120 (Amaya, *et al.*, 2004).

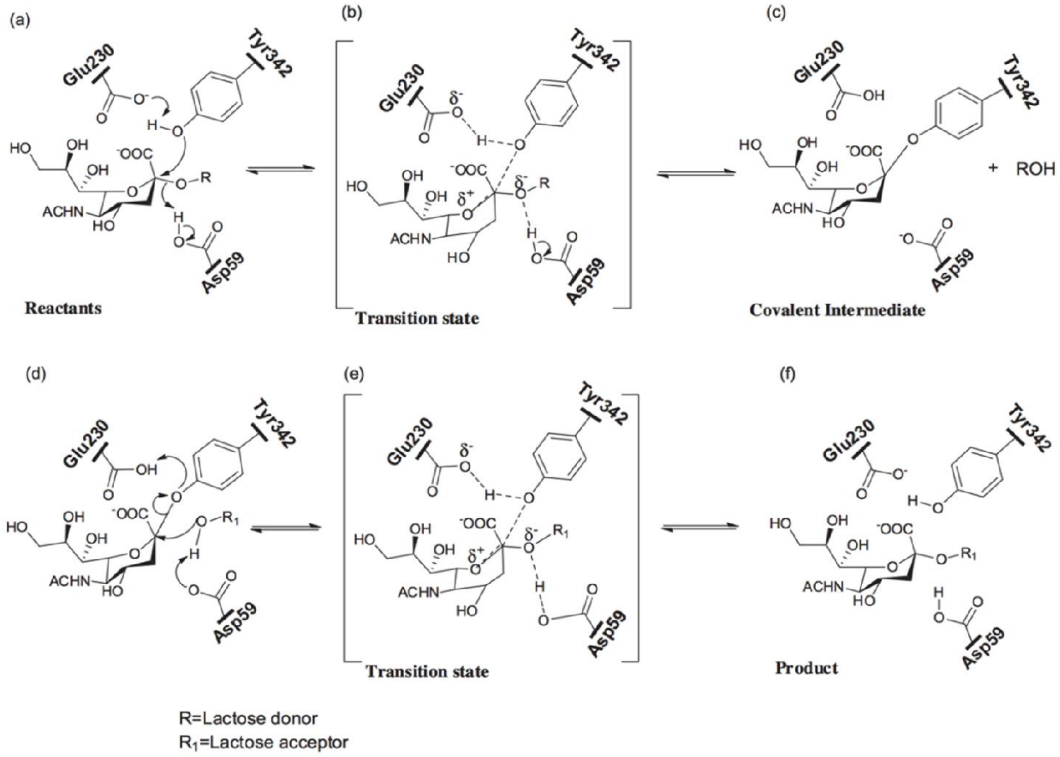


Figure 1.11. Catalytic mechanism of TcTS (Damager *et al*., 2008)

Tyr342 then attacks the C2 of the sialic acid in a S_N2-type reaction with the assistance of Glu230 (Buschiazzo *et al.*, 2000). These interactions immobilize the sialyl moiety in the sialic acid binding pocket. When SAL binds to the catalytic site, Asp59 is deprotonated on the hydroxyl group then provides a proton to the galactoside of cleaved glyosidic bond of donor substrate, followed by the release of the donor galactose induced by conformational changes of Tyr119 (Figure 1.12 C, D) (Amaya *et al.*, 2004).

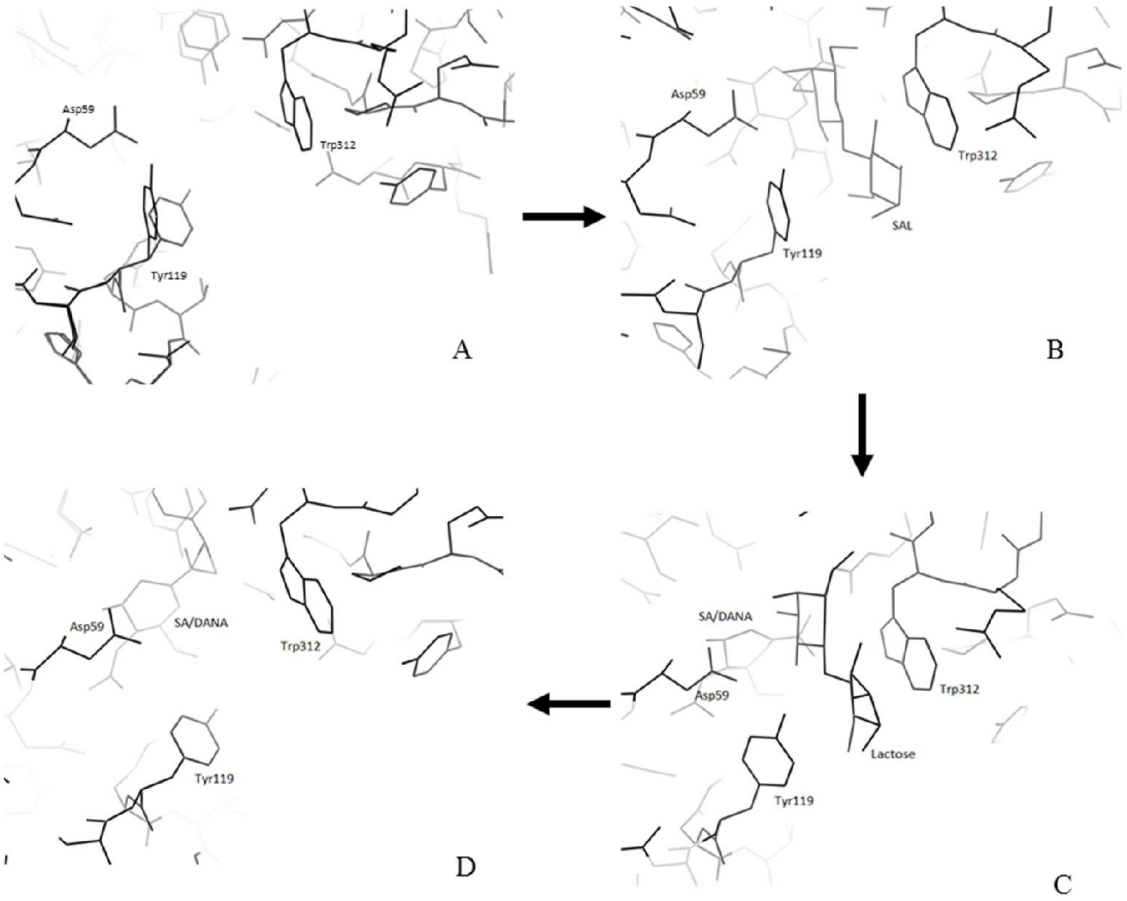


Figure 1.12. Arrangement of lactose binding site (Tyr119 and Trp312) during hydrolysis process: A. un-ligand TcTS catalytic site (PDB: 1MS3); B. Intermediate structure of TcTS complex with SAL (PDB: 1S0I); C. TcTS catalytic site complex with cleaved SA (DANA) and lactose (PDB: 1MS0). D. TcTS catalytic site complex with DANA (PDB: 1MS8) (Buschiazzo *et al.*, 2000, Amaya *et al.*, 2004).

The acceptor galactose enters the catalytic site only when the sialyl binding site is filled (Buschiazzolejandro *et al.*, 2002), the intermediate Asp59 then captures the proton from the acceptor galactose and triggers completion of the synthesis of SA-acceptor galactose complex (Amaya *et al.*, 2004).

The sialic acid binding site of TcTS is conserved with other GH family sialidases. Therefore, the unique sialylation is believed to be achieved by a gate formed by Tyr119 and Trp312 (Amaya *et al.*, 2004). Tyr119 is present in two different conformation in crystal structures of TcTS, either stacked parallel to Trp312 (closed conformation), or

flipped toward the sialic acid binding site (open conformation). In unliganded structures, Tyr119 can be in either of two conformation, indicating the flexibility of this residue in the absence of bound ligand (Figure 1.12 A). When the donor substrate SAL is bound to the active site, Tyr119 stays in the closed conformation and forms a sandwich stacking structure with the lactose moiety of SAL and Trp312 (Amaya *et al.*, 2004). After the hydrolysis process, Tyr119 flips to the open conformation, stabilizing the sialic acid binding by hydrogen bond formation and allowing the donor lactose to be released. In sialylation process, binding of the acceptor glycan triggers the conformational change of Tyr119 from open to closed conformation (Buschiazzo *et al.*, 2002). In published crystal structure studies, no conformational change of Trp312 has been identified. However, mutagenesis and molecular dynamic simulation (MDS) studies suggest that conformation changes of Trp312 also occur during the catalytic process and may be important for lactose binding and release of lactose (Mitchell *et al.*, 2010).

1.3. Development of TcTS inhibitors

1.3.1. Prelude

In the investigation of Chagas disease, TcTS has several characteristics of a potential drug target. Firstly, trans sialidase (TS) is unique to trypanosomes. Sialidase is a common enzyme in both prokaryotic and eukaryotic cells, however, trypanosome TcTS is the only type of enzyme which was found that has both hydrolase and transferase activity (Colli, 1993). This unique function of TcTS comes from its structure at the catalytic site which contains a lactose binding portion that can interact with donor or acceptor oligosaccharides. All the strong inhibitors of sialidases, only inhibit TcTS very weakly (Watts *et al.*, 2003). Additionally, TcTS is known as a key virulence factor in *T. cruzi* infection and critical for the survival of the parasite. *T. cruzi* has reduced ability to escape from immune surveillance without TcTS activity, caused by the lack of sialic

acid on its surface. The other functions of TcTS, such as assistance in cell invasion and the effect on the immune system, would also be removed by inactivation of TcTS. The importance and uniqueness make TcTS a potential drug target against Chagas disease. Also, the catalytic mechanisms and structure of TcTS has been intensively investigated during the last decade, these results are fundamental to structure-based rational drug design.

During the past three decades, many efforts were focused on identifying and developing new inhibitors of TcTS for applications in chemical therapy and biological investigation. As well as being a drug target, TcTS was also identified as a potential vaccination target, since it occurs in substantial quantity and is highly conserved on the *T. cruzi* surface.

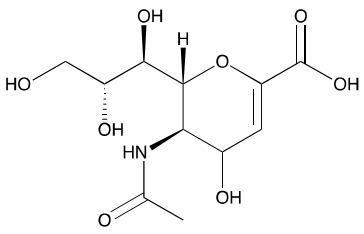
1.3.2. Sialic acid analogue inhibitors

SA, SA analogue or SA-like molecules were often tested as potential transition state inhibitors of TcTS. Large number of these molecules were good sialidase inhibitors, but the inhibition effect was significantly weaker for TcTS in most cases. For example, 2, 3-Dehydro-2-deoxy-N-acetylneuraminic acid (DANA) **(1)** is a good inhibitor against the influenza neuraminidase (K_i=4 μM) (Smith *et al.*, 2001), and *T. rangeli* sialidase (K_i=1.5 μM) (Paris *et al.*, 2005). However, it is only an extremely weak inhibitor against TcTS (K_i=12.3 mM) (Paris *et al.*, 2005).

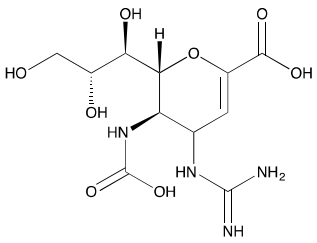
Unsurprisingly, other DANA derivatives, such as 4-guanidino-2,4-dideoxy-2,3-dehydro-N-acetylneuraminic acid (zanamivir) (**2**) and (3R, 4R, 5S)-4-acetamido-5-amino-3-(1-ethylpropoxy)-1-cyclohex-ene-1-carboxylic acid (oseltamivir) (**3**), which are both clinically approved drugs against influenza virus and inhibit sialidase at the nanomolar level, can only inhibit TcTS in very high concentration (K_i= 12.3 mM). 3-((1S)-1-(acetylamino)-2-ethylbutyl)-4- ((aminoimino-methyl) amino) - 2- hydroxy-(1S,2S,3R,4R)-cyclopentanecarboxylic acid (peramivir) (**4**) inhibits influenza neuraminidase in low nanomolar levels, but cannot inactive TcTS even at

concentrations up to 10 mM (McLaughlin *et al.*, 2015; Paris *et al.*, 2005). Another DANA analogue, 4-Amino-2-deoxy-2,3-didehydro-N-acetylneuraminic acid (**5**), which is also a successful influenza neuraminidase inhibitor, has no effect on TcTS (Kashif *et al.*, 2017).

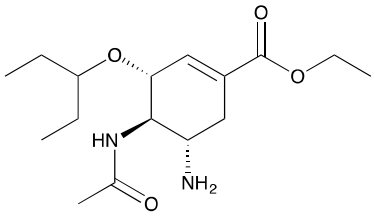
The 2,3-difluorosialic acid (DFSA) (**6**) has also been shown to be a weak TcTS inhibitor (K_i = 20mM) which shows time-dependent inhibition and covalent bond formation with Tyr342 (Watts *et al.*, 2003). As an irreversible inhibitor, intensive modification and screening of DFSA was carried out during the last decade. C7 modification of DFSA was performed to improve TcTS inhibition, the 7-N-anilino derivatives gave a IC_{50} of 3.2 mM (Resende, 2010). Modification at C9 was performed Withers group (**8**), this type of compound can also inhibit TcTS activity in a time-dependent manner and the introduction of lactose can result in reactivation (Buchini *et al.*, 2008). Modification at the glycerol side chain could be another strategy for TcTS inhibitor development, but further design and screening are still necessary. Therefore, traditional sialidase/ neuraminidase inhibitors have only a limited inhibition effect on TcTS, which possibly is a result of the unique binding site and catalytic mechanisms of this enzyme.



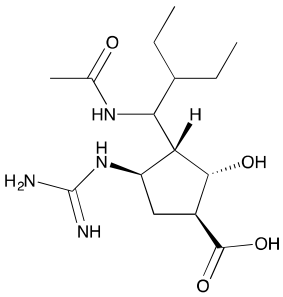
1. DANA



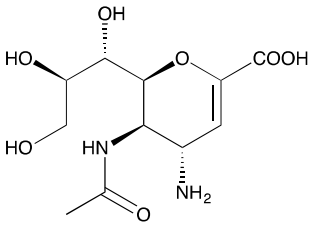
2. Zanamivir



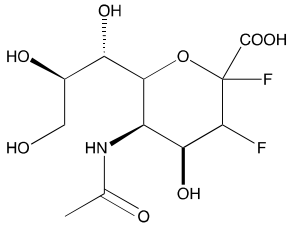
3. Oseltamivir



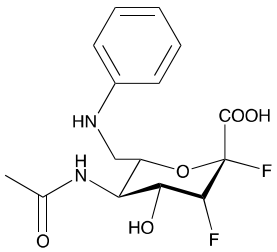
4. peramivir



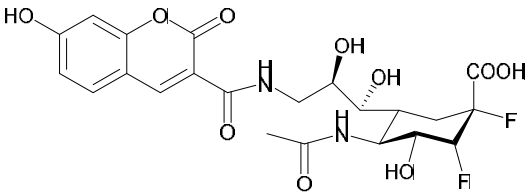
5



6



7



8

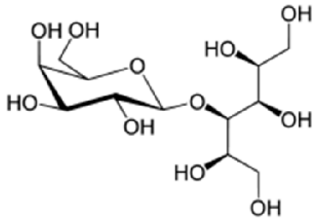
Figure 1.13. Structure of sialic acid analogue inhibitors (**1**: Smith *et al.*, 2001, **2-3**: McLaughlin *et al.*, 2015, **4:** Paris *et al.*, 2005, **5**: Kashif *et al.*, 2017, **6**: Watts *et al.*, 2003, **7**: Resende, 2010, **8**: Buchini *et al.*, 2008)

1.3.3. Glycoside analogue inhibitors

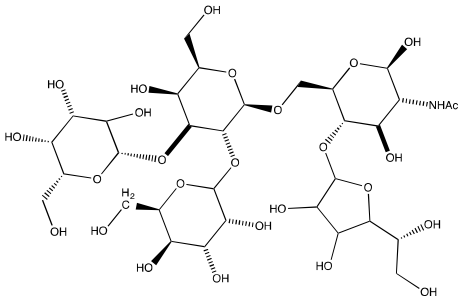
As a consequence of the unique transfer sialic acid activity, TcTS has developed an extra lactose binding site compared to sialidases. Thus, lactose and other glycans have also been considered as potential inhibitors against TcTS. TcTS has strong substrate specificity, some acceptor glycan molecules which can bind to the lactose binding site but only to be a poor sialic acid acceptor could be an inhibitor of TcTS. Lactitol (**9**) is the only commercial available TcTS inhibitor (Figure 1.13). As an acceptor analogue, which can bind to the catalytic site of TcTS with a K_m of 0.24mM, lactitol can inhibit TcTS *in vitro* (IC_{50}=0.57mM) and can reduce cell invasion of *T. cruzi* (Agusti *et al.*, 2004). Further research generating modifications on lactitol by adding Gal*p*, Gal*f*, or phenyl, failed to improve TcTS inhibition, the best compound Penta 5ol (**10**) gave an IC_{50} of 0.61mM. (Agusti *et al.*, 2007).

Substituted galactosides (**11**) were proved to reduce TcTS activity down to 37% but they are still a type of weak inhibitor and the best compound from the screen (**11**) has high cytotoxicity (Carvalho *et al.*, 2010). β-thiogalactopyranosides (**12, 13, 14**) are another type of glycoside which can inhibit TcTS activity with IC50s in the 0.5-1mM range (Harrison *et al.*, 2011).

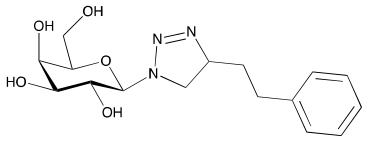
Another strategy of galactoside substitution is phosphorylation. Three types of phosphorylated galactose have been screened as TcTS inhibitors by Busse *et. al*. Compound **17**, the best inhibitor among their tested compounds, gives an IC_{50} of 1.5mM. However, other compounds (**15, 16, 18**) are within a range of IC_{50} from 3-8mM, suggesting that they cannot be potential tight binding TcTS inhibitors, but they could be compounds of interest for further development and potential inhibitors after conjugation with a sialoside (Busse *et al.*, 2007).



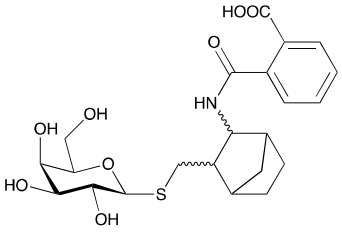
9. Lactitol



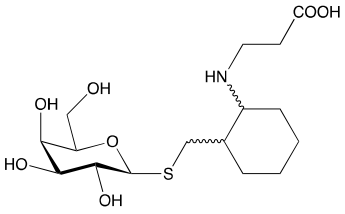
10



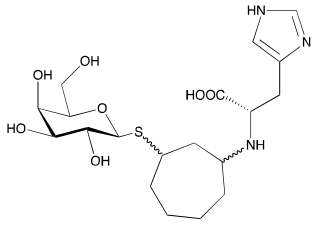
11



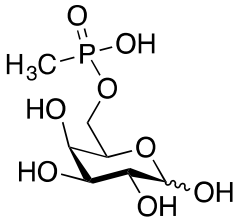
12



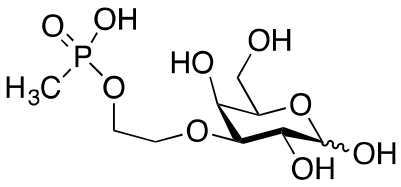
13



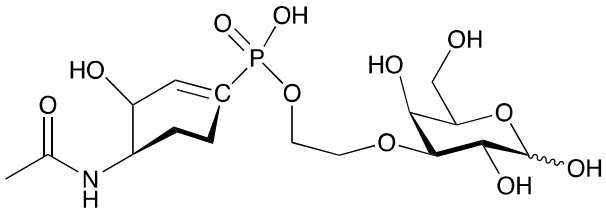
14



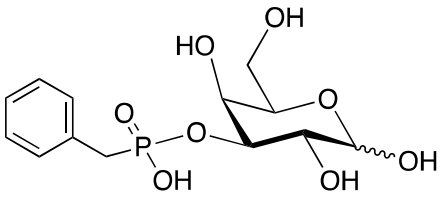
15



16

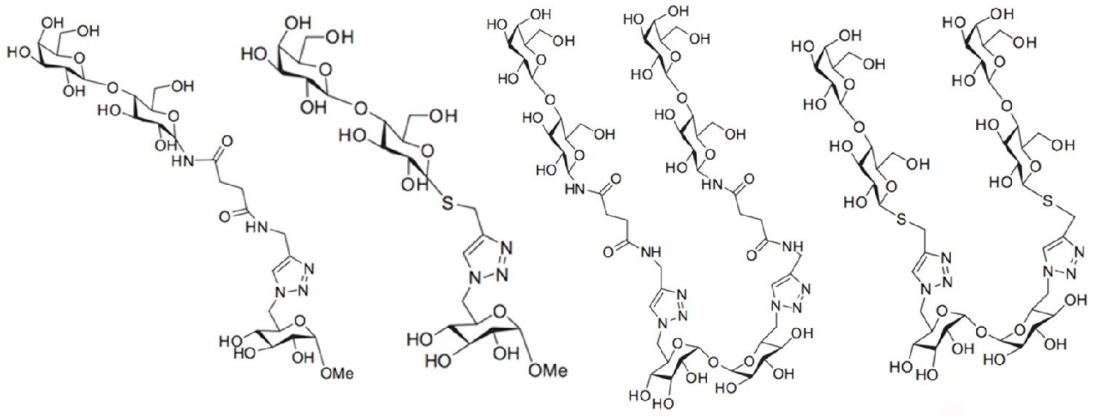


17

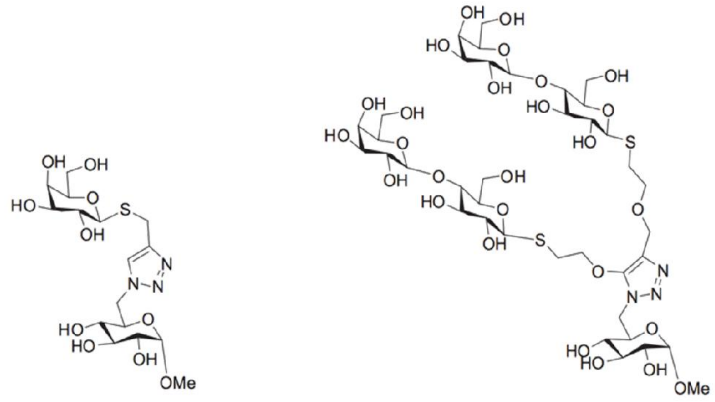


18

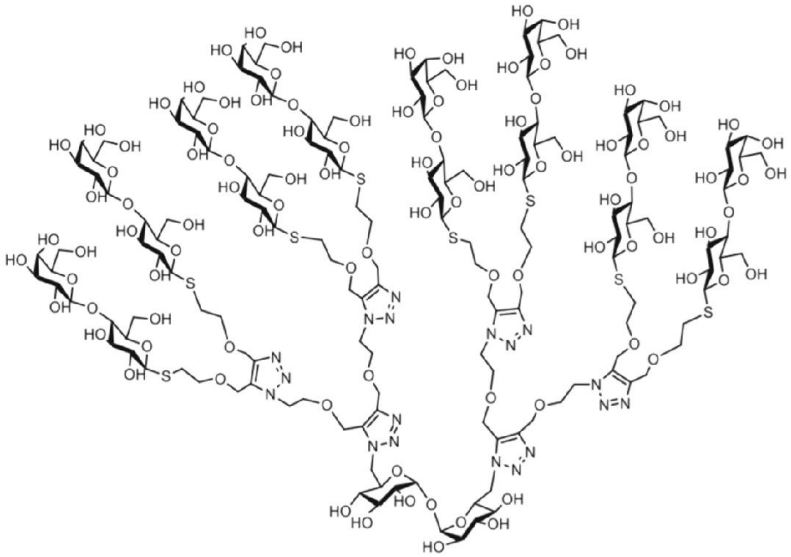
Figure 1.14. Structure of glycoside analogue inhibitors (**9**: Agusti *et al.*, 2004, **10**: Agusti *et al.*, 2007, **11**: *Carvalho et al., 2010,* **12-14**: Harrison *et al*., 2011, **15-18**: Busse *et al.*, 2007)



19 **20** **21** **22**



23 **24**



25

Figure 1.15. Structure of glycoside analogue inhibitors (**19-22**: Cano *et al.*, 2014, **23-25**: Agusti *et al.*, 2016)

In 2014, Cano *et. al* synthesized mono and divalent β-N- and β-S- galactopyranosides (Gal*p*), which are analogues of β-galactopyranosides (*β*Gal*p*), a terminal glycan and TcTS acceptor substrate on *T. cruzi* mucin. These βGalp analogues still can act as an acceptor substrate, but also reduce TcTS activity. The mono β-N-Gal*p* (**19**) and β-S-Gal*p* (**20**) play a major role as sialic acid acceptor and only can inhibit 32% and 41% of TcTS activity at 1mM respectively. The divalent analogues show better inhibition effect, which is 70% inhibition at 1mM for divalent-β-N-Gal*p* (**21**) and 53% for divalent-β-N-Gal*p* (**22**) (Cano *et al.*, 2014).

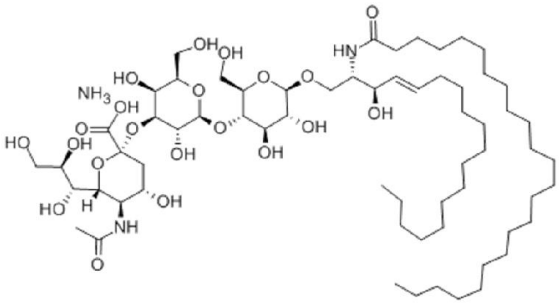
Recently, Cano and co-workers synthesized multivalent β-thiogalactopyranosides and thiolactosides. With same concentration of inhibitor, the mono glycocluster compound **23** shows 68% inhibition of TcTS whilst the divalent compound (**24**) have 58% inhibition, and the multivalent glycocluster (**25**) can inhibit 82% of TcTS activity (Agusti *et al.*, 2016). Most of the glycoside analogue inhibitors show competitive binding with acceptor substrate, and reduce the transfer activity, but none of these inhibitors has a greater effect than lactitol, and none of them can fully inhibit TcTS at the sub-millimolar level.

1.3.4. Sialyl-glycan analogue inhibitors

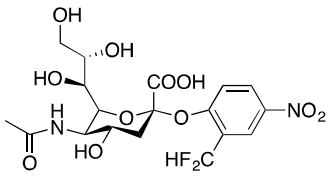
The unique dual-site catalytic pocket of TcTS resulted in the design of sialyl-glycan analogue inhibitors, a group of inhibitors which were designed to interact with both the sialic acid binding site and the lactose binding site. Ganglioside G_{M3} (**20**) is a natural substrate of TcTS, and it was found to inhibit TcTS activity (IC_{50}=10-100μM) (Vandekerckhove *et al.*, 1992). Later, several types of small molecule compounds were designed based on dual-site binding and evaluated as TcTS inhibitors. Carvalho *et al.* reported that 2-difluoromethyl-4- nitrophenyl-3,5-dideoxy-D-glycero-α-D-galacto-2-nonulopyranosid acid (NeuNAFNP, **21**) and its analogue dansyl-NeuNAcFP (**22**), can inhibit TcTS *in vitro* time-dependently and reduce cell invasion of *T. cruzi*. However, as a mechanism-based inhibitor, it can also inhibit sialidases and lost inhibition in

presence of a high concentration of substrates (Carvalho *et al.*, 2010). Neu5Aca2–3-Galb-O-octyl (**23**), another sialiyl disaccharide, has been proved to be a weak inhibitor that can reduce TcTS activity at the millimolar level (Harrison *et al.*, 2001).

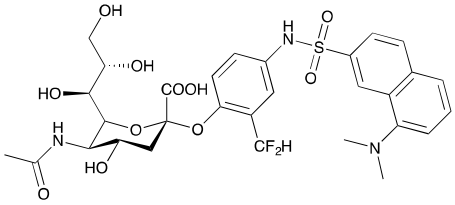
In most of the designs of sialyl-lactose based inhibitors, the structure of sialic acid is generally maintained to interact with the sialic acid binding site. By contrast, a recent design for a TcTS inhibitor used a molecular simplification of sialyllactose (SAL), keeping the full structure of the galactoside, replacing the glucoside with aryl groups and simplifing the sialic acid into a carboxyl group (**24**) or tetrazolyl group (**25**). However, 3-O-substituted-aryl-β-D-galactopyranosides inhibitors (**24, 25**) achieved only 20% inhibition at 1mM. Moreover, lack of inhibition-mode evaluation and the different binding manner of compound **24** and **25** given by docking evaluation, make it difficult to explain the inhibitory mechanism of this type of compound (Silva *et al.*, 2014).



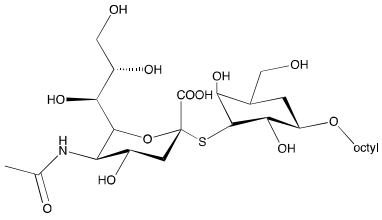
20. G_{M3} ganglioside



21



22



23

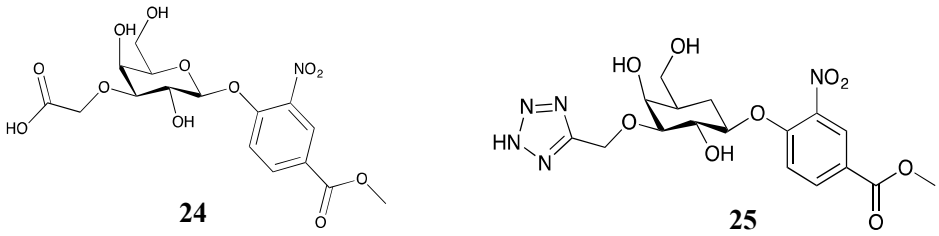


Figure 1.16. Structure of sialyl-glycan analogue inhibitors (**20**: Vandekerckhove *et al.*, 1992, **21-22**: Carvalho *et al.*, 2010, **23**: Harrison *et al.*, 2001, **24-25**)

Consequently, no strong inhibitor has been discovered from sialyl-glycan analogues. Most of the designs of this group of inhibitors keep the un-modified SA structure for SA pocket binding, and links the SA moiety and glycan moiety with a glycosidic bond, which can be cleaved by TcTS or other sialidases. Therefore, how to achieve a stable binding between inhibitor and TcTS without hydrolysis should be considered in further development.

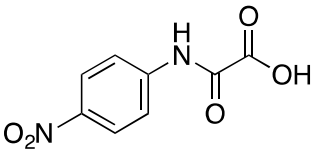
1.3.5. TcTS inhibitors not related to carbohydrate

As well as the inhibitors described above, some inhibitors have been described that are not related to carbohydrate structure. This group of inhibitors can be either molecules retrieved from database screening or known sialidase inhibitors, or molecules designed based on the structural architecture of TcTS catalytic site. Unlike sugar-based inhibitors, the mechanisms of these inhibitors are various, and many of them do not work as competitive inhibitors. However, so far, the best reported inhibitors have been these sugar independent molecules, and this group of inhibitors normally show high specificity against TcTS rather than cross-reacting with other sialidases.

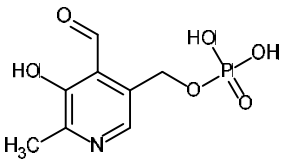
N-(4-Nitrophenyl)-oxamic acid (**26**, Figure 1.16) is a good sialidase inhibitor which can inhibit *Vibrio cholerae* neuraminidase in a non-competitive manner, but it can only inhibit TcTS weakly (Engstler *et al.*, 1994). Pyridoxal phosphate (**27**) has been proved to be a weak non-competitive inhibitor (K_i=7.3mM) when 4-Methylumbelliferyl-N-

acetyl-α-D-neuraminic acid (MuNANA) was applied as the substrate, and none of its analogues, such as pyridoxine and pyridoxamine phosphate were a stronger inhibitor (Ferrerogarcia *et al.*, 1993). Another type of phosphonate inhibitor, cyclohexenephosphonate derivatives, can inhibit TcTS in a similar range, the best compound of this group are N-acetyl-cyclohexenephosphonates monoalkyl esters (**28**, K_i=4.7 mM), but the inhibitory mechanism has not been evaluated (Streicher and Busse, 2006).

Benzonic acid has been proved to be an influenza neuraminidase inhibitor which interacted with the catalytic site (Chand *et al.*, 1997). Further structural based design of benzonic acid derivatives against TcTS has been performed by Neres *et. al*. (2007) Compounds were designed based on catalytic site architecture and inhibited TcTS activity successfully at sub-micromolar level (**29** and **30**, IC_{50} about 0.5mM). Furosemide (**31**), also known as Lasix, a drug already in clinical use for treatment of hypertension and edema, was unexpectedly found to inhibit TcTS with IC_{50} of 0.67mM, inhibiting TcTS by either non-competitive or mixed mechanisms, suggesting that benzonic derivatives cannot bind to the TcTS active site (Neres *et al.*, 2007). Later, Neres *et. al*. performed *in vitro* screening using a library containing 305,000 compounds based on the structural geometry of the TcTS catalytic site. The best candidate was 3-benzothiazol-2-yl-4-(2-isopropoxy) phenyl-but-3-enoic acid (**32**), which gave an IC_{50} of 0.15mM and 92% inhibition at 1mM. However, enzyme kinetics data showed that it is either a non-competitive or mixed inhibitor, indicating that a second inhibitor binding site exists (Neres *et al.*, 2009).

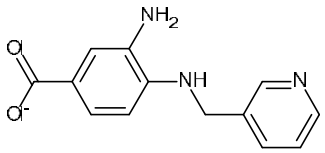


26

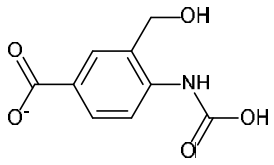


27

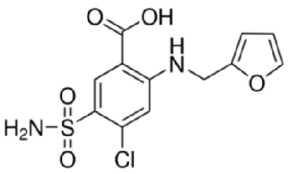
28



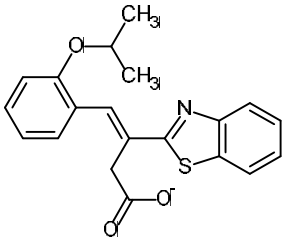
29



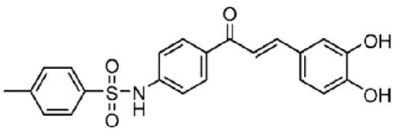
30



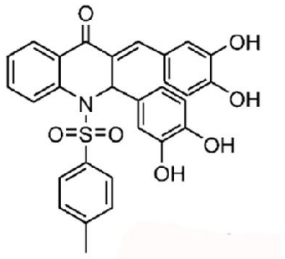
31



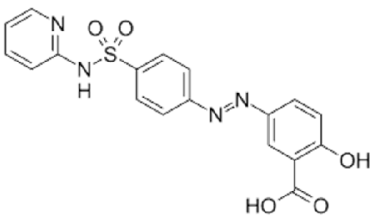
32



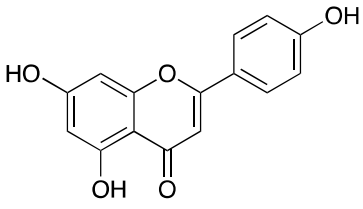
33



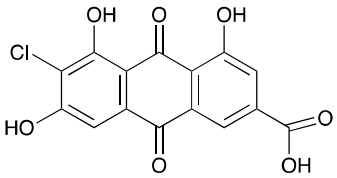
34



35 Sulfasalazine



36 Apigenin



37

Figure 1.17. Structure of non-sugar-related inhibitors (**26**: Engstler *et al.*, 1994, **27**: Ferrerogarcia *et al.*, 1993, **28**: Streicher and Busse, 2006, **29-31**: Neres *et al.*, 2007, **32**: Neres *et al.*, 2009, **33-34**: Kim *et al.*, 2009, **35**: Lara-Ramirez *et al.*, 2017, **36-37**: Arioka *et al.*, 2010)

Sulfonamide chalcone (SC) and derivatives were first designed as α-glucosidase inhibitors, and Kim *et.al* (2009) found that they can also inhibit TcTS. The 3, 4-dihydroxyl derivative (**33**) gave an IC50 of 0.9 μM. Another type of SC derivative, Quinolinone, also presents reasonable inhibition, tetra-hydroxy quinolinone (**34**) is one of the best known competitive inhibitors, with an IC_{50} of 0.6 μM (Kim *et al.*, 2009). Recently, *in vitro* screening based on the TcTS catalytic site structure using over 3000 FDA approved drug candidates has been carried out to find a TcTS inhibitor, sulfasalazine (**35**) which is structurally similar to sulfonamide chalcone, was found to show anti-trypanosoma activity (Lara-Ramirez *et al.*, 2017).

High throughput screening of TcTS inhibitors from a natural product library was performed by the Arioka and co-workers, and two scaffolds, flavonoid and anthraquinone were identified to have significant inhibition of TcTS. 4′,5,7-trihydroxyflavone (Apigenin, **36**) gave the minimum IC_{50} (78 μM) among flavonoid compounds, and 6-chloro-9,10-dihydro-4,5,7-trihydroxy-9,10-dioxo-2-anthracene-carboxylic acid (**37,** IC_{50}=0.58 μM**)** is the best potential inhibitor with a anthraquinone scaffold (Arioka *et al.*, 2010).

1.3.6. Summary

Following the discovery of this unique enzyme, many efforts have been made to develop an inhibitor against TcTS. However, so far there are still no strong, specific, and clinically safe inhibitors. Although some good candidates, such as lactitol, quinolinone and anthraquinone have been reported, all of them still have many clinically proven difficulties for profited use in therapy, further improvement and optimization in effect, specificity, cytotoxicity, and pharmacokinetics is required. Apart from the mechanism-based inhibitors, which are mostly weak inhibitors, many decent inhibitors are not related to TcTS substrate structure. Understanding the binding mechanism of these inhibitors will be helpful in further inhibitor design and improvement.

Since TcTS is not readily crystallized compared with many other sialidases, only two inhibitors, DANA (**1**) and compound **7**, have had their crystal structure in complex with TcTS published (Buchini *et al.*, 2008; Buschiazzo *et al.*, 2002). Both two inhibitors are sialic acid derivatives, which can be better recognized by TcTS and interact with the catalytic site. Some structural efforts have also been carried out for the substrate-unrelated inhibitors, but no crystal structure was able to be obtained by either co-crystallization or crystal soaking (Neres *et al.*, 2009). Knowing the binding site details and inhibition mechanisms are particularly important for the group of inhibitors that are unrelated to the TcTS substrate. Firstly, they are generally better inhibitors than sugar analogues and there is more potential for further development. Moreover, many of these inhibitors were designed based on catalytic site architecture but show either non-competitive or mixed inhibition (**29, 30, and 32**). Some compounds discovered by screening also tend to inhibit TcTS by interacting with other binding sites rather than catalytic site (**27, 36, and 37**).

1.4. Protein crystallography: history, principle and applications

In the investigation of biological and biochemical process, structural biology plays a pivotal role. The interaction details between the macromolecule acceptor (e.g. protein and nucleic acid) and the ligand, which could be small molecules or bio-macromolecules, can be interpreted visually by analysis of the structure of the molecules or complex. To interpret the structure of biomolecules, several methods have been developed, the two most commonly used high-resolution methods are X-ray crystallography (XRC) and nuclear magnetic resonance (NMR).

As a mature method, which has been developed and applied for 100 years, XRC has contributed 90% of the structures in the protein data bank (PDB). It can be used to analyse high resolution three-dimensional structures from small molecules up to macromolecules with a molecular weight of 10^6 g/mol such as proteins and nucleic

acids (Rupp, 2015). NMR is another major technique used to determine bio-macromolecule structure. Unlike XRC, NMR will provide 3D structure information in the solution state, and it can also yield dynamic information of protein or protein-ligand complexes. However, normally NMR can only be used for proteins smaller than 30 KDa (Cavanagh *et al.*, 1995). In the past decade, several techniques have been developed to improve the data quality in structure determination or break through the limitations of current techniques. For example, cryo-electron microscopy (Cryo-EM) is becoming a powerful technique for the determination of large molecular structures especially large proteins or protein complexes, which are not able to be crystallized and too big for NMR, and the highest resolution of Cryo-EM structures can reach 3Å (Bai *et al.*, 2015). Another technique applied in solution state is small-angle X-ray scattering (SAXS), which can detect structural parameters such as shape distribution for larger protein (>30KDa) (Blanchet and Svergun, 2013). For proteins for which large high quality crystal cannot be obtained, X-ray free electron laser diffraction (XFEL) or electron diffraction can also help to solve the structure.

The first protein crystal of earthworm hemoglobin was obtained in 1840s and several protein crystals were reported during the 19th century. In early years of the 20th century, protein crystallization was mainly applied as a technique for protein purification (McPherson and Gavira, 2014). The first protein crystal structure, sperm whale myoglobin, was determined by John Kendrew in the 1950s (Kendrew *et al.*, 1958), about 60 years after X-ray were discovered by Wilhelm Röntgen. Later, with the increasing importance of protein crystallography in studies of protein three-dimensional structure and function, number of new approaches have been developed in protein purification methods, crystal growth, improvement in data collection techniques and processing of diffraction data. Presently, X-ray crystallography is the most important technique in determination of protein and protein complex structures, whilst other methods such as NMR and SAXS are mainly used for dynamic studies.

Similar to many other molecules, proteins tend to form crystals in supersaturated aqueous solution (Rupp, 2010). However, the crystallization of macromolecules, including protein, nucleic acid or macromolecule complex, still has no comprehensive theory. The production of macromolecule crystals is based on several principles, previous experience and even an exhaustive search of conditions. The solution of macromolecule structure through XRC can be divided into several steps: (1). requirement of single crystal, (2) X-ray diffraction, (3) processing of X-ray data and (4) model building and refinement (Figure 1.18).

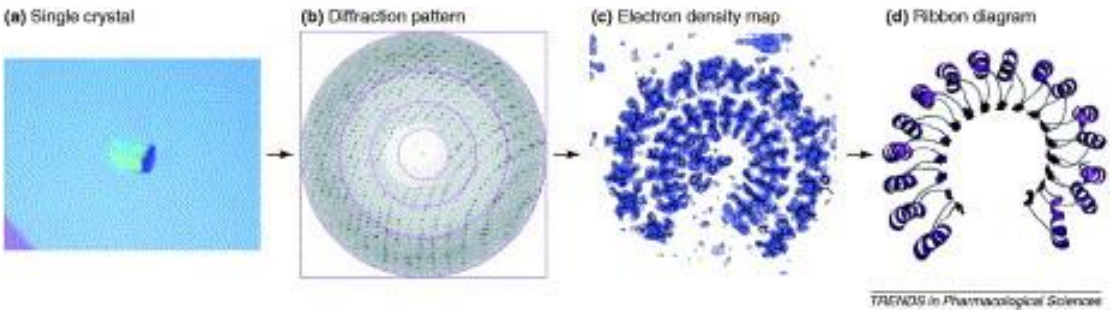


Figure 1.18. Brief procedure of protein XRC from crystal to structrure (Acharya and Lloyd, 2005)

XRC techniques rely on the exclusive scattering of X-rays, which is known as diffraction, by the electrons in the analyte molecule. X-ray diffraction from single crystals can be described by Bragg's law:

$$2dsin\theta = n\lambda$$

where **λ** is the wavelength of the X-ray, ***d*** is the distance between planes of identical repeating atoms in crystal, and θ is the angle of the incident light (Bragg and Bragg, 1913). The initial X-ray data collected is a set of diffraction images which reflect the periodicity of the crystal architecture, including the location and intensity of each set of planes in a crystal. The diffraction pattern contains diffraction spots, each of which correspond to a point of reciprocal lattice representing an X-ray wave with an amplitude (intensity of the light) and relative phase (a position of point of time). However, the detectors can measure the intensity of the signal but not the phases, which means the

overall scattering from a particular set of Bragg planes cannot be calculated, that is called "the phase problem". There several ways to solve the phase problem, for example, multiple isomorphous replacement (MIR), in which heavy atoms are inserted into the crystal then the scattering factors are calculated using the location of heavy atoms in the lattice which dominate the scattering (de La Fortelle and Bricogne, 1997). The phase problem can also be solved through molecular replacement (MR) if there is analogous structure with known phase (Scapin, 2013).

The structure factor F:

$$F_{(h,k,l)} = \sum_{j=1}^{atoms} f_{(j)} \exp\left[2\pi \cdot i(hx_{(j)} + ky_{(j)} + lz_{(j)})\right]$$

contains both the amplitudes "$f_{(j)}$" and phases "$\exp[2\pi \cdot i(hx_{(j)}+ky_{(j)}+lz_{(j)})$" of all the atom in the crystal, then the summation of these leads to an initial electron density map, through Fourier synthesis (Rupp, Bernhard, 2009):

$$\rho_{(x,y,z)} = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{(h,k,l)} \exp\left[-2\pi \cdot i(hx + ky + lz)\right]$$

.

Based on the electron density map, a model of the crystallized molecules can be built and refined. The structure factor used in calculation of the initial map, F_{obs} (or F_o), is only an approximation of the true F due to the phases being only an estimation. F_{calc} (or F_c) generated from the initial model is used to calculate the F_{obs}-F_{calc} map, also known as the difference map. During rebuilding, a $2F_{obs}$-F_{calc} map, which is calculated using both the observed structure factor and that calculated from the model, is normally used to inform the structural features and quality of the result in practice (Wlodawer *et al.*, 2008).

At the stage of initial phasing and subsequent model building, the model is still far from perfect. Therefore, additional rounds of refinement are required, to improve the phase and quality of the model. Refinement is normally achieved through statistical adjustment of the atomic coordinates so that they better fit the observed data. One of

the important parameters of the refinement process is the "working R factor" (R_{factor} or R_{work}):

$$R_{factor} = \sum_{hld} \left| |F_{obs}(hkl)| - |F_{cal}(hkl)| \right| / \sum_{hld} |F_{obs}(hkl)|$$

which describes the difference between the calculated structure amplitudes and the experimental amplitudes. However, approaches to reduce the R_{factor}, may introduce bias, because the refined model is changed to fit features of the electric density map, which might be incorrect or incomplete. Thus, the "free R factor" (R_{free}), which is calculated by the same methodology as the R_{factor} but using a section of data not used in the R_{factor} calculation, was introduced to validate the accuracy of the refined model (Brunger, 1992). Two methods are frequently used in refinement, maximum likelihood, which tries to minimize the R_{factor} (Murshudov *et al.*, 1997), and simulated annealing, which allows the introduction of additional randomness by heating then rerefining the coordinates while they are being "cooled down" (Brunger and Rice, 1997).

However, XRC also has some limitations. Firstly, large molecules are not easy to crystallize, the large size, high complexity and flexibility of biological macromolecules results in difficulty in crystallization. Secondly, the conditions under which a protein crystallises are normally unpredictable. Therefore, it can be a time-consuming process to find appropriate crystallization conditions. Additionally, it can be a problem to obtain the phases when direct methods cannot be used, for example, at lower resolution, when heavy atom soaking breaks the crystals, and no analogous structure can be used for MR. These several reasons restrict the application of XRC. Nevertheless, XRC is still one of the most reliable approaches in bio-macromolecule structure determination.

1.5. Summary of previous X-ray crystallography structure studies of TcTS

During the last decade, many efforts have been carried out to investigate the mechanisms and inhibition of TcTS using crystallographic methods. 10 crystal structures of TcTS have been published in PDB so far (Table **1.1**). The first crystal

structure of TcTS was determined by Buschiazzo *et. al.* in 2002, after introducing 7 surface mutations based on *T. rangeli* sialidase which has already been crystallized (Buschiazzo. *et al.*, 2002), including the unliganded structure in different crystal forms (PDB: 1MR5, 1MS3, 1MS4), the structure in complex with substrates (PDB code: 1MS0, 1MS5) and in complex with inhibitors (1MS1). These crystal structures showed the structure of TcTS, indicated the catalytic mechanisms which differ from other sialidases, and also presented the secondary binding site for lactose constituted by Tyr119 and Trp312.

Table 1.1 List of the crystal structures of TcTS in the PDB

PDB code	Structure details	Reference
1MS0	Monoclinic form of TcTS, in complex with DANA and lactose	Buschiazzo *et. al.* 2002
1MS1	Monoclinic form of TcTS, in complex with DANA	
1MS3	Monoclinic form of TcTS	
1MS4	Triclinic form of TcTS	
1MR5	Orthorhombic form of TcTS	
1S0I	TcTS in complex with sialyl-lactose	Amaya *et al.*, 2004
1S0J	TcTS in complex with MuNANA	
2AH2	TcTS in complex with 2,3-difluorosialic acid	
3B69	TcTS complex with benzoylated NANA derivative (inhibitor)	Buchini *et al.*, 2008
3POZ	TcTS in complex with the Fab fragment of a neutralizing monoclonal IgG antibody	Buschiazzo *et. al.* 2012

However, the first series of TcTS structures failed to obtain a structure with the full-length substrate, neither the natural substrate SAL nor the non-natural substrate MuNANA. This was primarily because after binding to active TcTS, even within a

crystal, the substrate can still be hydrolysed and sialyated in the presence of an appropriate acceptor, therefore, when soaking or co-crystallizing active TcTS with substrate it is difficult to obtain a catalytic intermediate structure with high occupancy level of substrate. In addition, conformational changes of TcTS during the catalytic process might damage the crystal, leading to poor diffraction. To solve this problem, Amaya et.al introduced an extra mutation D59A in the catalytic pocket, removing the negative side chain of residue Asp59 which breaks the substrate glycosidic bond, thus the natural ligand can bind stably without hydrolysis (Amaya *et al.*, 2004). The Michaelis complex structure of mutated TcTS and substrate is the first structure in which both SA moiety and lactose moiety were present in the catalytic site. Even though the construct using in this study is inactive, the Michaelis complex structures (Figure **1.19**) provide strong evidence of the catalytic mechanisms, especially the mechanism of transfer SA activity.

Crystal structure studies are also widely used in drug discovery. So far, only limited numbers of drug-bound TcTS structures have been reported since there is still no effective inhibitor for TcTS. As a weak inhibitor (IC_{50}=12mM), DANA is the most intensively studied inhibitor against TcTS. Based on both crystal soaking and co-crystallized structures, DANA binds to the SA binding site of the TcTS active site and has similar interaction with TcTS compared to those seen with SA. Moreover, when a TcTS crystal was soaked with NA-S-Gal (PDB: 1MS05) only DANA could be defined in the catalytic site instead of SA, and no galactose could be seen in the structure (Buschiazzo. *et al.*, 2002). These results suggest that DANA can be a by-product of the catalytic activity and explains why it is only a weak inhibitor of TcTS. Although it binds well to the active site, it is unable to form the glycosidic bond due to its lack of the C2-oxygen. Another two NANA derivatives, 2,3-difluoroSA and 5-(acetylamino)-9-(benzoylamino)-3,5,9-trideoxy-3-fluoro-D-erythro-α-L-manno-non-2-ulopyranosonic acid (BFN), were also study by XRC (Buchini *et al.*, 2008). All these ligand-soaked structures were of TcTS in complex with SA derivatives inhibitors, however, neither a soaked or co-crystallized structure has been obtained with the best competitive inhibitor,

although molecular dynamic simulations have been predicted the interaction between inhibitor and TcTS catalytic site. Interestingly, a large number of SA derivatives have been reported to be good binders in the active site due to their high occupancy in crystal structure, but they are only weak inhibitors of TcTS.

However, some of the best reported inhibitors are non-competitive inhibitors, which means they do not bind in the catalytic site of TcTS but in another unknown binding site (Arioka *et al.*, 2010). Apart from the catalytic pocket, no other binding site has been defined for TcTS, not even in the lectin-like domain, which has been shown to bind carbohydrates in many sialidases and *T. congolense* trans-sialidases (Waespy *et al.*, 2015). Therefore, interpreting the structure of TcTS complex with non-competitive inhibitors will be helpful to understand more details of the protein-ligand interaction and the multiple functions of TcTS.

1.6. Overview of this thesis

The aim of this PhD project was to design, synthesis and evaluate α-aminophosphonate compounds as a new type of TcTS inhibitor which was designed by structure mimicry of the natural substrate of TcTS.

Chapter 2 describes the design and the synthesis of α-aminophosphonates derivatives. The expected binding manner of these α-aminophosphonates is competitive binding with both the SA binding site and lactose binding site, and the substituents were designed based on a stable interaction with the catalytic site. The structural geometry and binding energy have also been predicted by molecular docking. In the second step, α-aminophosphonate compounds were synthesized through a one-pot microwave assisted reaction followed the deprotection step. 22 compounds were synthesized and screened.

Chapter 3 discusses the structure-activity relationship (SAR) of α-aminophosphonates. The best compound shows good inhibition and high selectivity compared with the previously-reported TcTS inhibitors. Nevertheless, after determination of the inhibition mode, all compounds were shown to inhibit in a non-competitive manner, suggesting that α-aminophosphonate does not interact with the TcTS active site.

Chapter 4 describes the crystal structures of TcTS complexes with inhibitors. Crystal structures of TcTS-α-aminophosphonate complexes were determined to identify the novel binding site between the catalytic and the lectin-like domain, which is the first reported allosteric site of TcTS. The novel binding site was validated by site-direct mutagenesis and introduction of anthraquinone, a known non-competitive inhibition by itself, into the α-aminophosphonate compound.

Chapter 5 provides further investigation of the mechanism of the allosteric inhibition by α-aminophosphonate, dual soaking with inhibitor compounds and lactose has been measured and lactose was detected in lactose binding site, which is not visible when crystals are soaked in lactose alone. From the crystallization results hypothesis was drawn that allosteric binding of α-aminophosphonate compounds induce the formation of lactose binding site, which is normally induced by SA binding, then blocks the entrance or release of substrates, resulting in inactivation.

All results are generally discussed with a conclusion in **Chapter 6**. The experimental details are given in **Chapter 7**.

The identification of the inter-domain allosteric site provides a new structural site for further rational drug design and suggests ideas for investigation of the catalytic mechanisms. This study favors further TcTS inhibitor development as a potential treatment for Chagas disease.

Chapter 2. Design and synthesis of α-aminophosphonate derivatives

2.1. Prelude

TcTS has been considered as a potential target for therapy of Chagas disease. However, the unique structure of the catalytic site compared to other sialidases results in traditional sialidase inhibitor only presenting weak inhibition against TcTS. Therefore, it is urgent to develop novel inhibitors based on the uniqueness of TcTS, not only to develop drug candidates against Chagas disease, but also to benefit further investigation of TcTS mechanism and functions. However, the broad screening of compounds using databases has been done without any strong inhibitor being identified. The design of new compounds could be a potential approach to discover potential TcTS inhibitors.

In modern drug discovery and development, structure-based drug design plays an increasingly important role as better structure determination techniques have increased the quality of target-ligand complex structures. Structure-based drug design has been successfully applied in development of influenza neuraminidase inhibitors. In terms of TcTS, as an enzyme which has unique structure and catalytic mechanism within the sialidase/neuraminidase family, there is the potential to develop specific inhibitors through investigating the protein-ligand structure.

This chapter describes the structure-based design of α-aminophosphonate derivatives as effective and selective inhibitors against TcTS activity. Initially, the α-aminophosphonate scaffold was designed based on structure mimicry of sialyllactose (SAL), the natural substrate of TcTS. Molecular docking also shows that the α-aminophosphonate compounds bind to TcTS catalytic site with a high predicted binding affinity, as designed. α-aminophosphonate compounds are synthesized by a two-step reaction.

2.2. Design of α-aminophosphonate derivatives as TcTS inhibitors

Structure-based drug design is regarded as a useful, rational, and promising method in drug development. In the past few decades, the improvement of protein structure determination techniques such as X-ray crystallography and NMR makes it possible to design specific new drugs via analyzing the structural characteristics of specific targets. Although there are still limitations of structure-based drug design, for example candidate compounds might not interact with the predicted binding site even though calculations predict good docked geometry and binding energy, there are still some success stories. For example, the discovery of zanamivir, an inhibitor toward influenza neuraminidase (which also targets sialic acid), was based on the crystal structure of influenza neuraminidase, and the structural characteristics of the active site were also considered in the inhibitor's further development. Thus, for discovery of novel inhibitors targeting TcTS, structure-based drug design could be a useful strategy to find drug candidates which are specific to TcTS.

The natural substrate of sialidase family is sialyl-glycans (Palese *et al.*, 1973). However, TcTS only shows hydrolysis and transfer activity for 3'- sialyl rather than 6'-sialyl, and shows highly effective activity on 3'-Gal substrates (Vandekerckhove *et al.*, 1992). Therefore, a common substrate, α-NeuNAc-(2→3)-β-D-Gal-(1→4)-D-Glc (sialyllactose, SAL, Figure 2.1 A) was selected as a model for rational design of potential TcTS inhibitors by structure simplification and mimicry.

Based on the interaction with TcTS, the structure of SAL can be divided into four functional regions for rational drug design: lactose, sialic acid, sialic acid carboxyl and glycosidic oxygen (Figure 2.1 A). Firstly, the lactose (or other oligosaccharides) moiety which stacks within the lactose binding site, and the sialic acid moiety which interacts with the sialic acid binding pocket which is structurally conserved among the sialidase/trans-sialidase family. Part of the sialic acid moiety, the carboxyl group interacts with the Arginine triad (Arg35, Arg245, Arg314). The glycosidic oxygen that

links the sialic acid and lactose moiety through a glycosidic bond, is cleaved by TcTS through Asp59, which acts as an acid catalyst (Amaya, *et al.*, 2004).

In the design of inhibitors, to mimic the structure of the lactose, the cyclic monosaccharide structure was simplified and replaced by aromatic groups (R_2 group, Figure 2.1 B), since rigid aromatic groups have more stable structures than the monosaccharide cyclic structure and might have stronger π-stacking with TcTS Tyr119 and Trp312 (Figure 2.1 C) (Hunter and Sanders, 1990). Potential R_2 groups could be benzene, naphthalene, biaryls, or anthracene (Figure 2.1E).

For the sialic acid moiety, the cyclic configuration of sialic acid was replaced by other cyclic structures, for example alkyl, phenyl, C_{5-6}heteroaryl, C_{5-6}carbocyclyl, or C_{5-6}heterocyclyl, with suitable substituents (R^X) (Figure 2.1 D). The carboxyl group is another important region for consideration for rational drug design for both sialidase and trans-sialidase due to the highly-conserved arginine triad. In most cases of rational drug design based on sialic acid structure, the carboxyl group tends to remain, but also there are examples where it is replaced by other moieties, and a phosphate group is one of the most common substitutes (Wang *et al.*, 2016; White *et al.*, 1995). In the research reported here, a phosphonate group is also being used to replace the carboxyl group.

Replacement of the carboxyl group in SA by a phosphate has been used in the development of sialidase/neuraminidase inhibitors. For example, the phosphonate analogue of SA is an influenza neuraminidase inhibitor (**38**, K_i= 0.2mM, Figure 2.2) (Chan *et al.*, 1997). Replacing the carboxyl group in peramivir to phosphonate also improves the inhibition effect against influenza NA **(39**, IC_{50} = 5.2 nM, Figure 2.2) (Wang *et al.*, 2016).

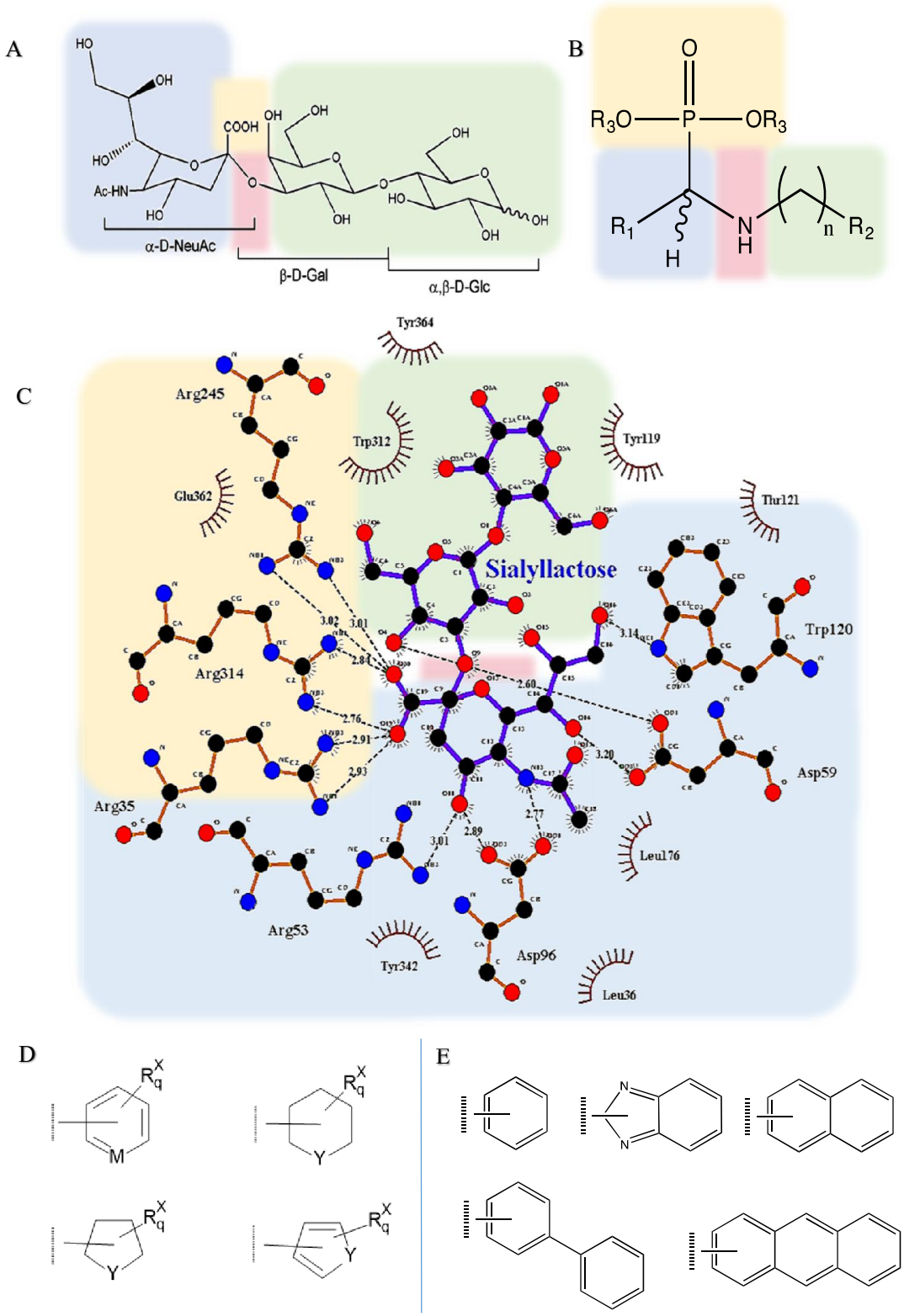
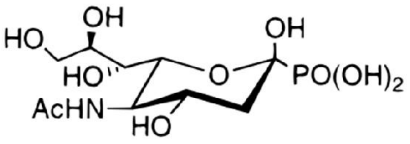
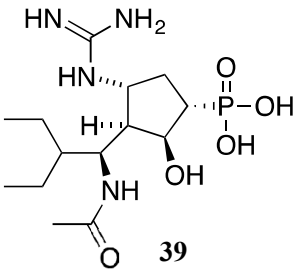


Figure 2.1. Structure of Sialyllactose: Yellow: carboxyl group of SA; blue: rest of SA, green: lactose and red: glycosidic oxygen (A), design of α-aminophosphonate inhibitor by molecular mimicry (B), and the interaction between SAL and TcTS, as seen in PDB: 1S0I, diagram generated by LigPlot Plus (C) (adapted from Amaya, *et al.,* 2004). Structures of potential R_1 groups (D) and R_2 groups (E)



38



39

Figure 2.2. Structure of phosphonate analogue of SA (38) and peramivir (39)

Therefore, in this design of TcTS inhibitors, the cyclic structures of SA, lactose and glucose are replaced by aromatic groups, which are expected to bind to the lactose binding site through π- π interaction. The oxygen at the glycosidic bond is changed to be a nitrogen, which is expected to interact with Asp59 stably via acid-base interaction, rather than being hydrolyzed.

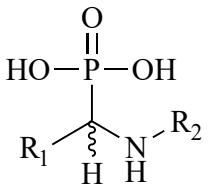
2.3. Prediction of α-aminophosphonate-TcTS interaction by Molecular docking

The α-aminophosphonate compounds were designed to inhibit TcTS activity by interacting in the catalytic pocket. To predict the position and affinity of TcTS-inhibitor interactions, molecular docking methods were used to simulate the interactions at the TcTS sialyllactose binding pocket. Protein models were prepared based on the structure of the active state of TcTS (PDB: 1S0I). Firstly, the screening of different R_2 groups was performed with the sialic acid mimic fixed as benzene. From the prediction by Autodock Vina (Table 2.1), the R_2 groups with the strongest binding were the 1-Naphthalene and 1-Anthracene. Therefore, these two substituents were selected for sialic acid moiety virtual screening. The ring of the sialic acid was replaced by a phenyl moiety with substituents including nitro, acetamido, and carboxyl groups.

As the results of the docking position show, (figure 2.3), the phosphate was predicted

to have appropriate binding with the arginine triad with interaction distances between 2 and 3 Å. The amino nitrogen and N-linked aryl group are also at the predicted position. The introduction of naphthalene or anthracene did not seem to change the overall binding conformation. In terms of the substitutions at the sialyl-mimic phenyl, 4-acetamido and 4-nitro groups interacted with Asp96 and Trp120 respectively (Figure 2.3 B).

Table 2.1. Predicted binding energy generated by Autodock Vina (Trott, *et al.*, 2010)



Compound	R_1	Residue predicted to bind R_1	R_2	Predicted affinity with TcTS[a] (kcal/mol)
41	Phenyl	N/A	Benzene	-4.5
42			1-Naphthalene	-8.6
43			2-Naphthalene	-6.8
44			2-Biaryl	-4.8
45			3-Biaryl	-4.6
46			2-Methyl-benzodiazol	-5.3
47			1-Methyl-naphthalene	-4.9
48			1-anthracene	-10.3
49	4-methoxyphenyl	Asp96	1-Naphthalene	-8.9
50	4-chloride phenyl	N/A		-8.8

51	4-trifluoromethyl phenyl	N/A		-7.5
52	3,5-bis-trifluoromethyl phenyl	N/A		-8.2
53	4-acetamido phenyl	Asp96		-9.6
54	4-nitro phenyl	Trp120		-9.8
55	4-nitro phenyl	Trp120	2-Biaryl	-8.7
56	4-nitro phenyl	Trp120	3-Biaryl	-7.9
57	4-trifluoromethyl phenyl	N/A	1-anthracene	-9.9
58	3,5-bis-trifluoromethyl phenyl	N/A		-10.9
59	4-acetamido phenyl	Asp96		-12.3
60	4-methyl ester phenyl	Asp96		-11.8
61	4-benzonic acid	Trp120		-12.1

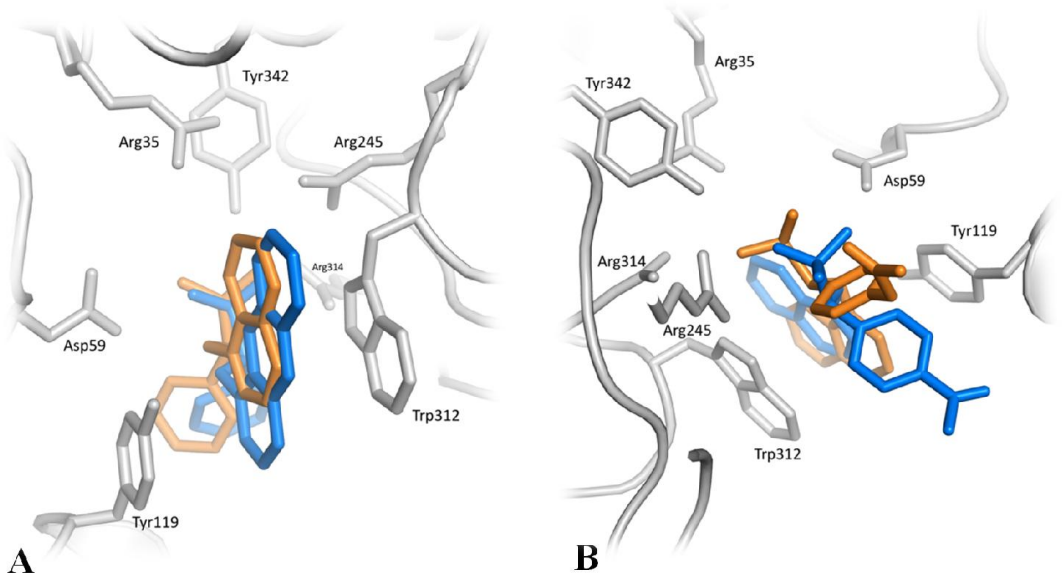


Figure 2.3. Docking conformation of different predicted compounds with TcTS. (A). Binding conformation of compound with R_1=H, R_2=naphthalene (orange) or anthracene (blue), respectively. (B). Docking of compound with R_2= naphthalene, R_1= acetamido (orange) or nitro (blue) respectively.

2.3.1. Synthesis of α-aminophosphonate compounds

α-aminophosphonate compounds have been shown to have several chemical properties: firstly, they are known as effective chelating agents (Kiss *et al.*, 1994), they can also inhibit nucleation and crystal growth of small molecules, for example barium sulfate and calcium carbonate (Guo and Severtson, 2004; Jones *et al.*, 2003), furthermore, they present high stability in harsh chemical environment.

Applications of aminophosphonates can be found from agriculture to medicine; enzyme inhibitors, anticancer agents, antibiotics, neuro-modulators, plant growth regulators and herbicides, antibacterial, and many other uses (Flynn *et al.*, 1985).
Aminophosphonate compounds play an important role in medical chemistry since it contains a P-C bond and amino bond (Figure 2.4). During last fifty years, plenty of α-aminophosphonate compounds have been synthesis and intensively studied as the analogues of amino acids (Naydenova *et al.*, 2010).

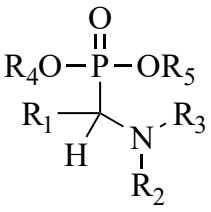


Figure 2.4. Structure of α-aminophosphonate

The chemical synthesis of aminophosphonate has been intensively studied and numbers of methods are now available. Organophosphorus compounds, which are widely used in organic synthesis, are reported to be used in aminoalkanephosphonic acids synthesis. However, the most widely used and the earliest reported method for the synthesis of amino-phosphonates is the Kabachnik-Fields reaction (KFR), which prepares an α-aminophosphonic acid using a carbonyl compound, an amine and a dialkyl phosphonate as starting reagent (Fields, 1952; Kabachnik and Medved, 1954). The mechanism of

KFR can be divided into two steps (Figure 2.4). The first step is the reaction of an aldehyde with an amine to form an imine, and the second step is the reaction of P-H bond from phosphonate with the C=N bond in the imine (Pudovik reaction) to form the aminophosphonate (Naydenova *et al.*, 2010).

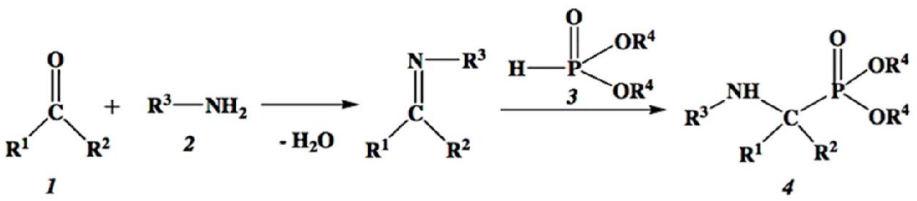


Figure 2.5. Mechanism of one-pot KFR (Naydenova *et al.*, 2010)

The reaction methods and conditions were intensively studied. In our study, two conditions were tested: microwave heating (solvent free, 90W, 5min) and reflux (THF, 50°C, 24 hours). The result (Figure 3, MPharm. Thesis, Chang, 2012) presents that solvent free microwave heating obtained significantly higher yields than long reflux, thus the subsequent reactions followed the one-pot solvent free microwave method.

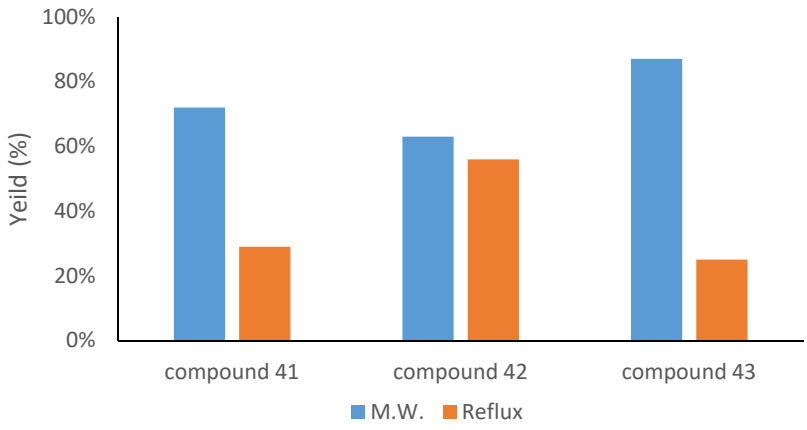


Figure 2.6. Yield of one-pot KFR through microwave heating and Reflux (adapted from Chang, 2012)

An aldehyde and amine (1:1 in molarity equivalent) are desolved in dimethyl phosphite (5-10 molarity equivalent). The reaction mixture is heated for 5min at 90W in a kitchen microwave. After reaction, the reaction mixture is extracted with CH_2Cl_2 and washed with ddH_2O. The organic phase was dried over anhydrous $MgSO_4$, filtered and

concentrated. The mixture was chromatographed on silica gel (EtOAc/PetEther) to give the protected compound.

In the second step of reaction, the methyl groups in dimethyl phosphite were removed using trimethylsilyl bromide (TMSBr) (Figure 2.6). One equivalent (molarity) of protected compound was dissolved in anhydrous CH_2Cl_2 (2ml) and anhydrous DMF (1ml), following resting on an ice bath for 10min, the TMSBr was added dropwise (2.5-5 equivalent). The reaction mixture was stirred on the ice bath for 30min, then was allowed to warm up to room temperature for 2.5hr to 24hr. The reaction was monitored by detecting the chemical shift of ^{31}P-NMR, which will completely shift from 25ppm to around 16ppm. ddH_2O is added to the reaction mixture and the content is neutralized to pH8 by the addition of 1M NaOH dropwise. The product was concentrated and purified on a C-18 reverse phase column (MeOH/H_2O).

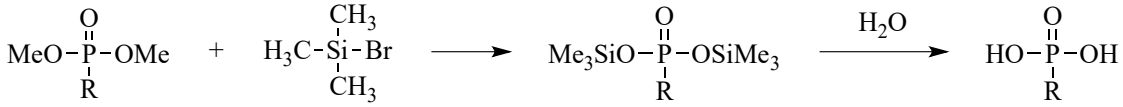
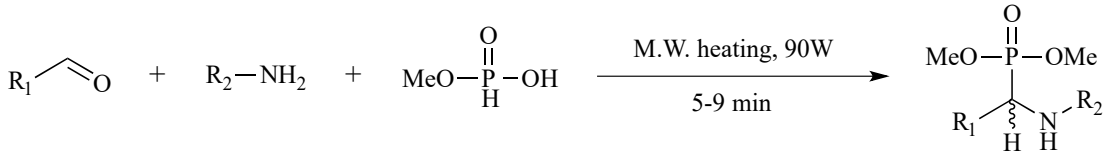


Figure 2.7. Mechanism of deprotection by TMSBr (Gutierrez, *et.al.,* 1999)

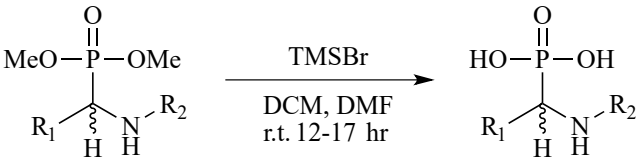
α-Aminophosphonate compound **41-61** have been synthesized for SAR screening (compound **43, 46, 47, 49** and **56** have been synthesized by MSc student Poyee Kowk). The reaction condition and yield of first step microwave assist KFRs are shown in table 3.2. Conditions and yield of deprotection reaction is listed in table 2.3 (detailed spectroscopic data are shown in Chapter 6.2).

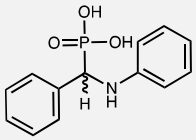
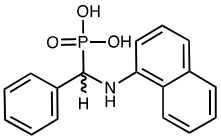
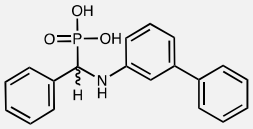
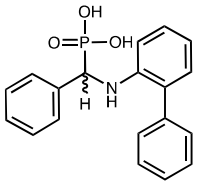
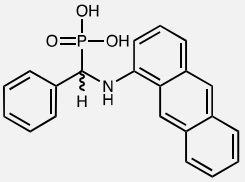
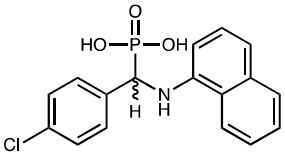
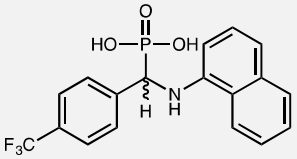
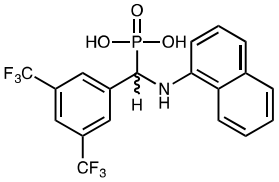
Table 2.2. List of protected aminophosphonate compounds

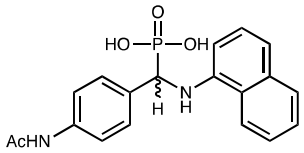
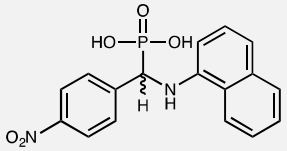
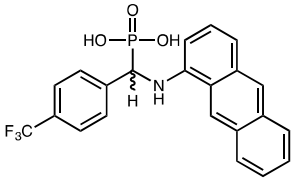
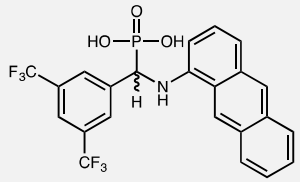
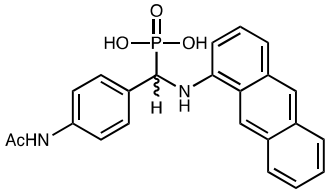
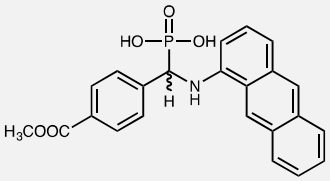
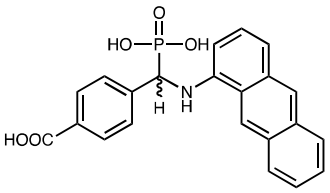


Compound	*R_1*	*R_2*	*Time (min)*	*Yield (%)*
S41	Phenyl	Phenyl	5	75
S42	Phenyl	1-Naphthalene	5	85
S44	Phenyl	2-Biaryl	7	84
S45	Phenyl	3-Biaryl	5	76
S48	Phenyl	1-Anthracene	9	87
S50	4-Chloridephenyl	1-Naphthalene	5	87
S51	4-Trifluoromethyl phenyl	1-Naphthalene	7	55
S52	3,5-Bis-trifluoromethyl phenyl	1-Naphthalene	7	84
S53	4-Acetamido phenyl	1-Naphthalene	8	72
S54	4-Nitro phenyl	1-Naphthalene	6	72
S57	4-Trifluoromethyl phenyl	1-Anthracene	8	21
S58	3,5-Bis-trifluoromethyl phenyl	1-Anthracene	7	76
S59	4-acetamido phenyl	1-Anthracene	8	43
S60	4-methyl ester phenyl	1-Anthracene	7	72

Table 2.3. Yield of deprotected aminophosphonic acid



Compound	*Starting compound*	*Equivlent of TMSBr*	*Product*	*Yield (%)*
41	S41	5		94
42	S42	3		82
44	S44	4.5		74
45	S45	7.5		86
48	S48	8		74
50	S50	4.5		79
51	S51	5		64
52	S52	3.5		40

Compound	*Starting compound*	*Equivlent of TMSBr*	*Product*	*Yield (%)*
53	S53	2.5		94
54	S54	4.5		59
57	S57	5		56
58	S58	3		63.1
59	S59	4		41
60	S60	6.5		76
61*	60	-		80

*Compound **61** has an extra de-protection step of compound **60**

Generally, the KFR step gives averagely good yield about 70 %, and the yield of naphthalene/bi-aryl compounds (mainly higher than 70%) are better than the anthracene compounds (40 % - 50 %). That might result from the fact that the anthracene compounds are more difficult to purify, i.e. compound **57** and **58** require extra column chromatography purification to obtain the desirable purity (>95 %), and further purified by recrystallization was performed in the synthesis of compounds **59** and **60** to remove the impurities. These extra purification steps could significantly reduce the yield of the products.

The second step gave the desired compounds which were designed as TcTS inhibitors, with a yield from 40 % - 80 %. The deprotected compounds have good solubility, which is appropriate for a TcTS inhibitor that will function extra cellularly, since TcTS is tethered to the outside of the Tc membrane. And avoiding penetration across the host cell membrane can also reduce the potential for cytotoxicity, because within the cell anthracene can lead to DNA damage (Bolognesi *et al.*, 1991).

Chapter 3. Evaluation of α-aminophosphonate as TcTS inhibitor

3.1. Prelude

As described before (Chapter 2), initial structure α-aminophosphonate derivatives were designed through molecular mimicry of SAL, and further modifications are generated based on the interaction between SAL and TcTS catalytic site and virtual screening of a series of compounds performed by molecular docking. In computational screening, the α-aminophosphonate structure fits the geometry of TcTS catalytic site, and the binding energy predicted that the α-aminophosphonates could be potential class of TcTS inhibitor with the lowest docking score (predicted binding energy) of about −12 kcal/mol, which is equal to micromolar levels of half maximal (50%) inhibitory concentration (IC_{50}) (Cosconati *et al.*, 2010).

This chapter will begin with the description of the IC_{50} screening of synthesized α-aminophosphonate compounds (**41-61**) to evaluate inhibition effects of this novel type of inhibitor and whether the results of *in vitro* screening can be related to the virtual screening. The specificity of this type of potential TcTS inhibitor through comparing inhibition effects against bacterial sialidases will also be described. Furthermore, the inhibition mode will be discussed and inhibition constant (K_i) will be calculated. Later, time-dependent inhibition and thermodynamic studies were carried out to explain the mechanism of TcTS inhibition by α-aminophosphonates.

3.2. Expression, purification and characterization of TcTS

Screening of expression conditions of TcTS had been carried out previously (Chen, MSc thesis, 2013) with a range of IPTG from 0.2mM to 1mM, and of temperature from 16°C to 37°C. Induction at 25°C using 0.5mM IPTG was found to be the optimal condition . In His-trap nickel affinity chromatography, a single peak was present in the early stage of elution (around 20%-50% of elution buffer). In the example shown, peak fractions A4-A7 were collected (Figure 3.1 A). Peak fractions were evaluated by SDS-

PAGE (Figure 3.1 B), the darkest band for each fraction at around 71 kDa is the predicted product. However some impurity bands are still present at around 50, 45, 30 and 25 kDa. The fractions that eluted from the His-trap column gave 75-85% purity which is adequate for enzyme activity determination or other low-sensitive enzyme-ligand assays, but did not reach the purity needed for crystallization or high-sensitivity assays.

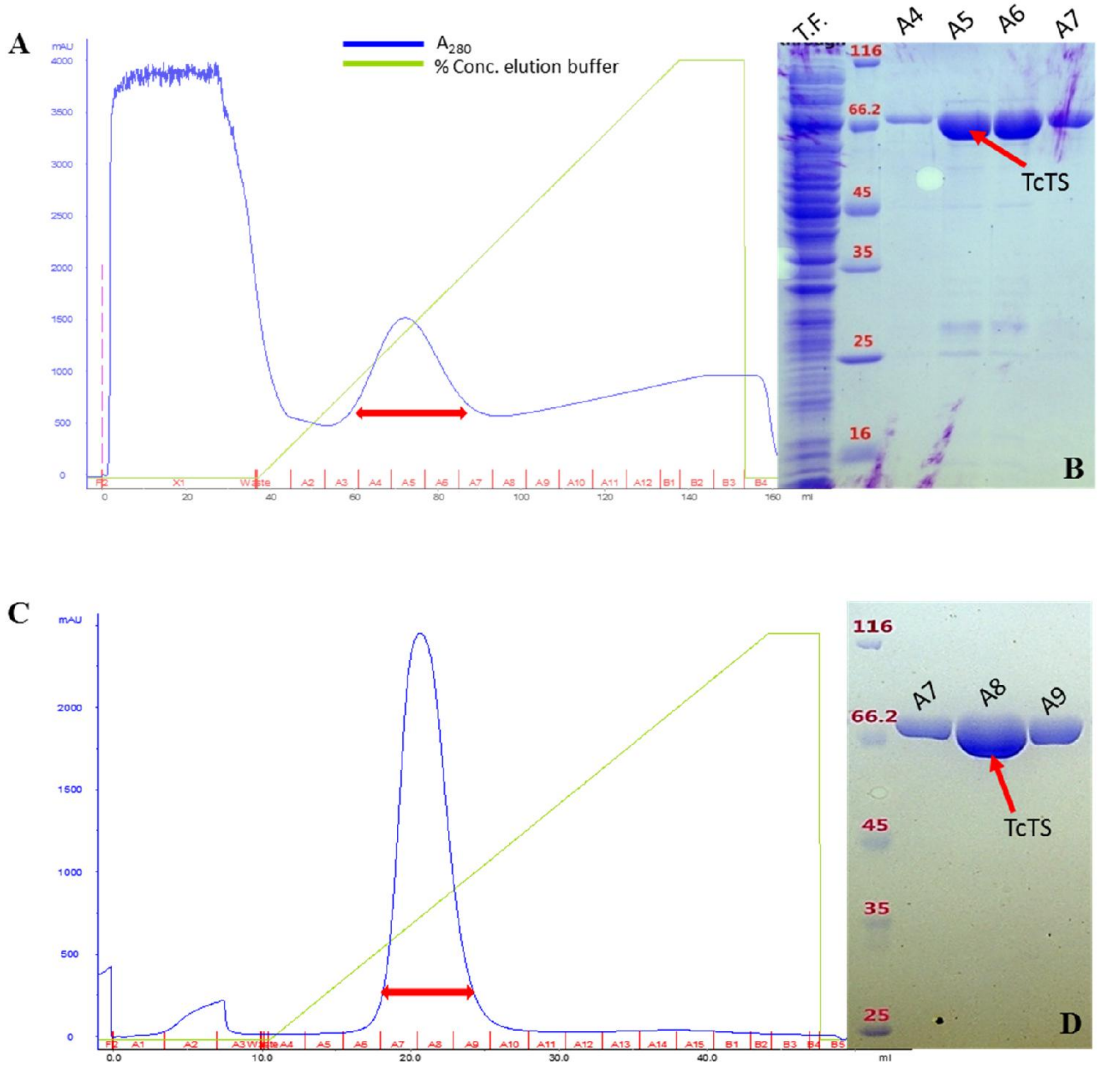


Figure 3.1. Chromatogram of TcTS His-trap purification (A) and SDS gel of peak fractions (B); Chromatogram of TcTS Anion-exchange purification (C) and SDS-PAGE gel of peak fractions (D).

Anion exchange purification was applied to the sample collected from His-trap

purification (fraction A4-A7) to eliminate the impurities. An example chromatogram showing peaks eluted from the anion exchange column is shown in Figure 3.1 C, from which fractions A7 to A9 were collected. As shown on the SDS-PAGE gel (Figure 3.1 D), all three fractions show a single band which is of the size of TcTS (71kDa), showing that the secondary purification has reached a level of purity that is adequate for crystallization. 3-5mg of protein was obtained from a 1mL culture. Pure TcTS was stored in PBS with 0.2% BSA, pH7.4, and placed at -80 °C until use.

Measurement of enzyme kinetics is important for drug candidate screening. Although the enzyme kinetics parameters of these enzymes are known, it was worthwhile to determine these before inhibitor screening since differences may occur in different environments. The Michaelis constant K_m, maximum activity rate Vmax, and turnover number k_{cat} were determined.

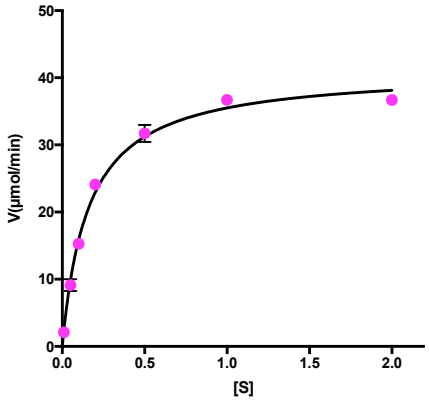


Figure 3.2. Non-linear curve fitting of rate against substrate concentration assay of TcTS

Table 3.1. Kinetic parameters for the hydrolysis of MuNANA by the TcTS

	$K_m(mM)$	k_{cat} (s^{-1})	k_{cat}/K_m $(mM^{-1}\ s^{-1})$
TcTS	0.15±0.03	3.8±0.43	25.3

The result of kinetics characterization is shown in Figure 3.2 and table 3.1. The K_m value of TcTS hydrolysis using MuNANA as substrate is about 0.15mM, which is similar to that in the literature (0.19 mM, Amino *et al.*, 1998), and the turnover number (k_{cat}) is 3.8 S^{-1}.

3.3. The identification of α-aminophosphonate as TcTS inhibitor

The assay to determine TcTS activity and the inhibition effect of the novel inhibitors used 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid (MuNANA), which has been widely used for detecting the enzyme activity of sialidase and TcTS. In this case, MuNANA was employed as the substrate of TcTS, the enzyme activity and inhibition being evaluated through fluorescence at the wavelength of 360 nm and 460 nm from released 4-methylumbelliferone via catalytic reaction.

The first step of catalytic activity is to cleave sialic acid from host sialyl oligosaccharides. Therefore, determining the hydrolysis activity in the presence of inhibitors is the quickest way to evaluate a potential inhibitor, and it has been the most common strategy in candidate screening of TcTS (Arioka *et al.*, 2010). Structure-activity (SAR) screening of α-aminophosphonates against TcTS hydrolysis activity began by looking at different substitutions in R_2 groups, described as the lactose mimicry group. Based on the R_2 group screening, the best candidates were selected for further screening of R_1 groups, the sialic acid binding substitutions. The IC_{50} assays were performed using purified TcTS with the concentration 0.05 mg/ml. TcTS in assay buffer ($TcTS_{assay}$, Appendix 8.4) was initially pre-incubated at 35°C for 10 mins with 10 µl inhibitor (dissolved in distilled water) in a gradient of final concentration from 15mM to 0.001mM. By detecting the changes florescence levels of reaction mixture with different concentration of inhibitor in the same period of time, the percentage of inhibition and IC_{50} can be calculated.

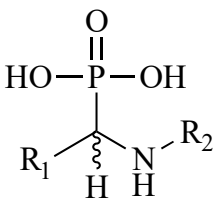


Figure 3.3 Basic skeleton for SAR screening of α-aminophosphonate derivatives

SAR analysis of R_2 groups

SAR screening of R_2 group sought to optimize the aromatic group that stacks in the lactose binding site. This screen substituted R_2 for different aromatic groups from a single phenyl to multiple ring structures such as naphthalene and anthracene whilst maintaining the sialic acid mimicry moiety as a phenyl group. The IC_{50} results of compounds with different R_2 group including phenyl, napthyl, biaryl and anthracene, are shown in Figure 3.4 and Table 3.2.

Firstly, the inhibition effect is positively correlated to the number of aromatic centres. Compound **41** (R_2=phenyl) cannot fully inhibit TcTS activity even in high concentration (40% activity remaining in 10mM compound 41), but the inhibition effect is significantly increased for the compounds which have multiple ring aromatic groups in R_2, such as naphthalene and anthracene. Additionally, the conformation of the R_2 groups also has a significant effect on the inhibition effectiveness. For instance, compound **42** which has 1-napthalene in R_2 gives an IC_{50} of 0.47mM, whilst the compound that has 2- naphthalene in the same place (compound **43**) has an IC_{50} of only 0.75mM. Similar results are also observed for biaryl compounds, the 2-biaryl compound (compound **44**) has a significant lower IC_{50} then the 3-biaryl compound (compound **45**). Furthermore, an additional methyl group between the amino group and the aromatic group does not help to improve the inhibition effect against TcTS, for example, both compounds **46** and **47** only show limited inhibitory effect.

R_2 group screening indicates that α-aminophosphonate derivatives are possible candidates to inhibit the activity of TcTS. The results of molecular docking suggested that the type and conformation of R_2 group are related to the predicted binding energy, which is roughly borne out by the SAR screening. Remarkably, several aminophosphonate compounds present IC_{50} values in low millimolar to high micromolar levels, indicating that α-aminophosphonate derivatives could be a potential of inhibitor against TcTS. Compound **42** (IC_{50}=0.43mM) and compound **48** (IC_{50}

=0.041mM), with naphthalene and anthracene in R_2 group, were highlighted in this first round of screening and were used for further screening of R_1 groups.

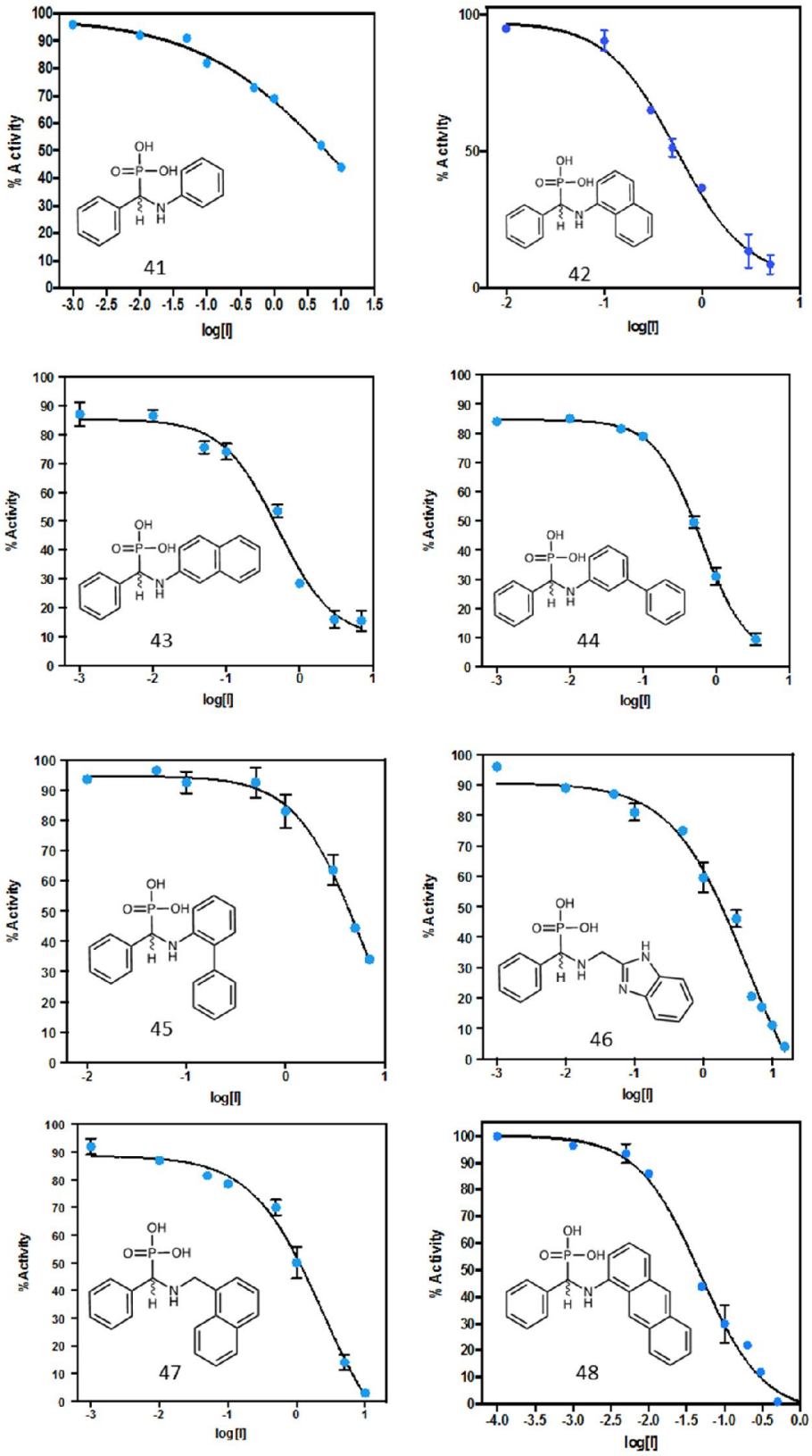
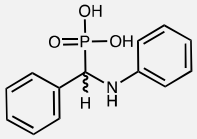
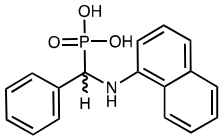
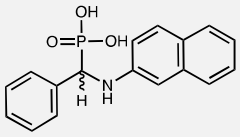
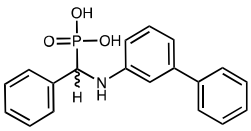
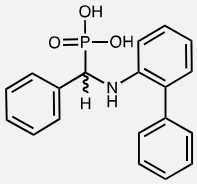
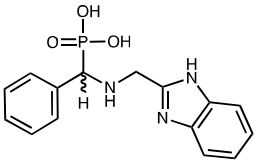
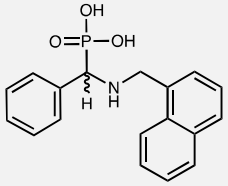
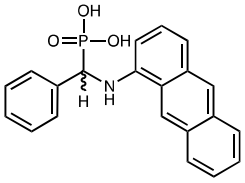


Figure 3.4 Non-linear curve fitting of R_2 group screening

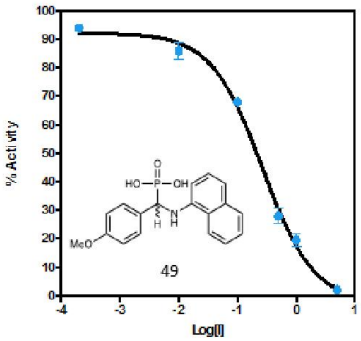
Table 3.2. List of α-aminophosphonate compounds and their IC_{50} against TcTS for R_2 screening

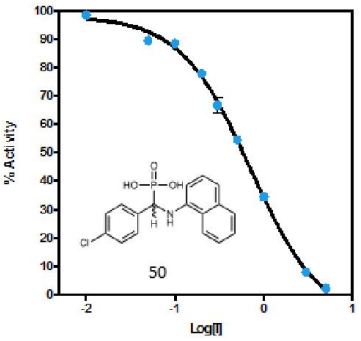
SAR analysis of R_1 groups

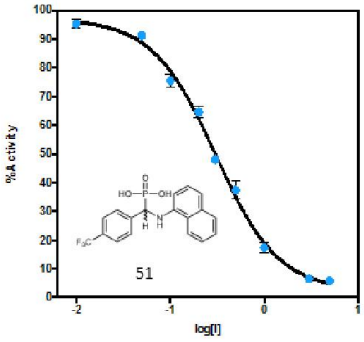
Compound	*Structure*	*Log(IC_{50})*	*St. error LogIC50*	*IC_{50} (mM)*
41		~0.888	~1.190	~7.73
42		-0.321	-0.073	0.470
43		-0.295	0.076	0.51
44		-0.204	0.025	0.63
45		0.746	0.240	5.58
46		0.621	0.199	4.24
47		0.375	0.209	2.374
48		-1.318	0.045	0.041

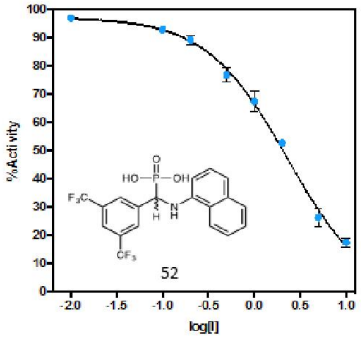
Based on the result of R_2 screening, several highlighted candidates were selected for the screening of the R_1 groups, including 1-naphthalene and 1-anthracene, which are two of the R_2 groups that have the highest inhibitory effect. Bi-aryl groups are also used as R_2 in to evaluate the R_1 group in the presence of different R_2 groups. The R_1 group screening was carried out to improve the inhibition effect, and possibly, specificity by changing various substitutents on the sialic acid moiety phenyl. In this study, the alternation of R_1 groups in the sialic acid moiety is achieved by changes to different modified benzaldehydes in the chemical synthesis. The R_1 groups that have been screened in this study include nitro, methyl, acetamido, carboxyl, halogens and trifluoromethyl. All the screened R_1 groups have been evaluated through molecular docking for their ability to interact with critical residues in the TcTS catalytic site whilst still fitting the structural geometry.

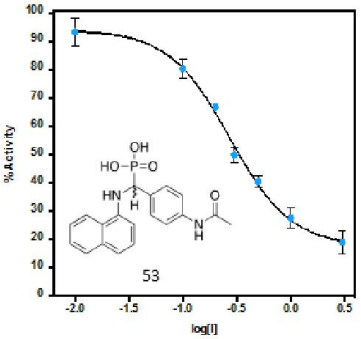
The IC_{50}s resulting from the R_1 group screening are shown in figure 3.5 and table 3.3. The addition to the R_1 group in the sialic acid moiety phenyl group can either increase, maintain or decrease the inhibitory effect against TcTS. As a side chain on sialic acid, introduction of an acetamido group in the sialic acid moiety can significantly help aminophosphonate compounds to inhibit TcTS activity. The IC_{50} of the 4'-acetamido-phenyl-naphthalene compound (compound **53**) is 0.23 mM, which is about half that of the phenyl-naphthalene compound (0.47 mM, compound **42**), the IC_{50} of which is 0.47 mM. The same effect of an acetamido group has been shown in anthracene compounds, compound **59**, the 4'-acetamido-phenyl-anthracene compound is the best inhibitor in this study, which has an IC_{50} of 15 μM, significantly lower than the related phenyl-anthracene compound, compound **48** (IC_{50} = 45 μM). The addition of a carboxyl group at the 4' position of the sialic acid moiety phenyl group obtains a decent drop of IC_{50} (compound **61**, IC_{50} = 21 μM). A 4'-Nitro group was shown to improve inhibition of naphthalene and bi-aryl compounds (compound **14, 15** and **16**), but synthesis of thenitro-anthracene compound failed due to difficulties in purification.

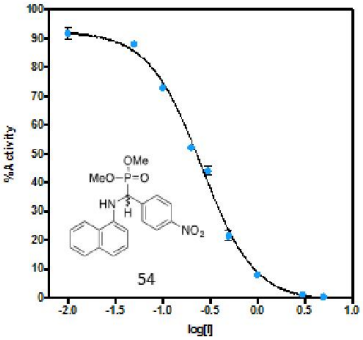

% Activity
Log[I]
49

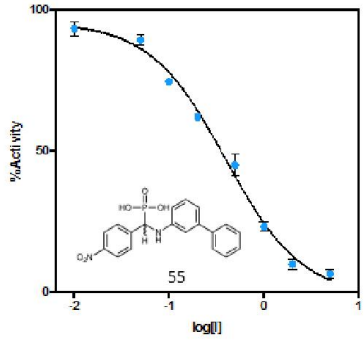

% Activity
Log[I]
50

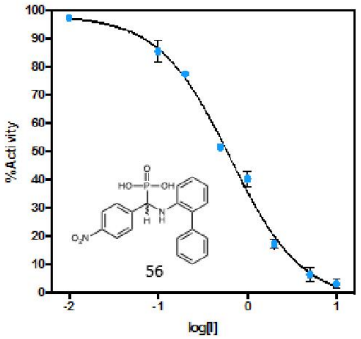

%Activity
log[I]
51

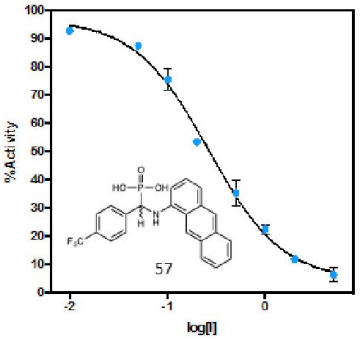

%Activity
log[I]
52

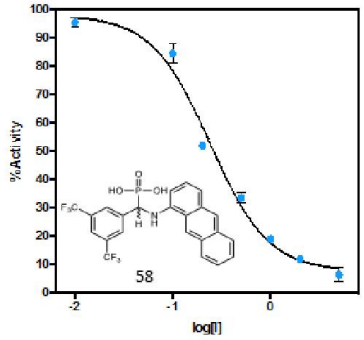

%Activity
log[I]
53

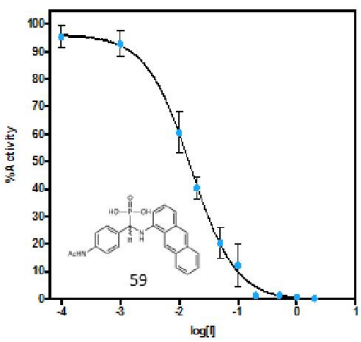

%Activity
log[I]
54

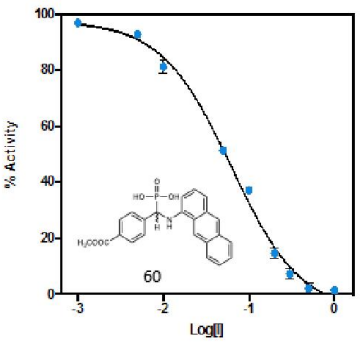

%Activity
log[I]
55


%Activity
log[I]
56


%Activity
log[I]
57


%Activity
log[I]
58


%Activity
log[I]
59


% Activity
Log[I]
60

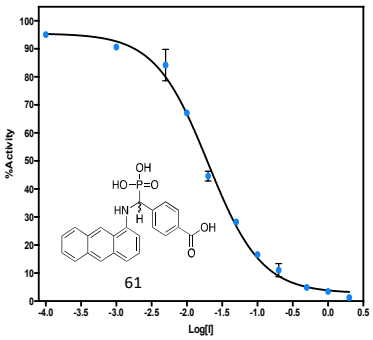
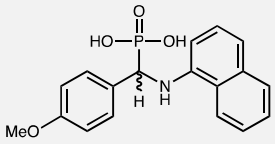
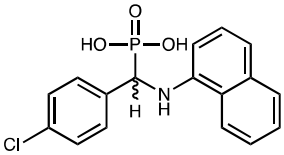
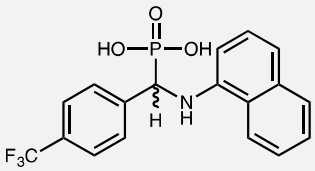
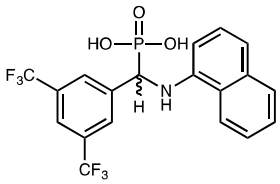
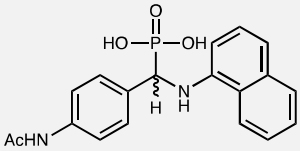
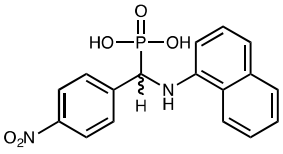
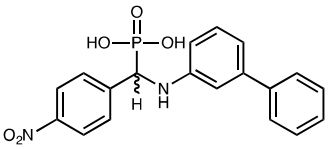
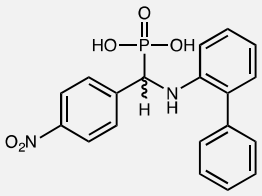
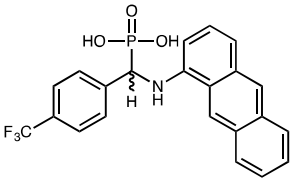
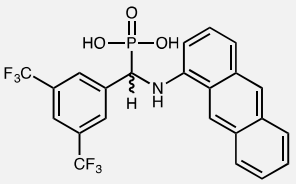
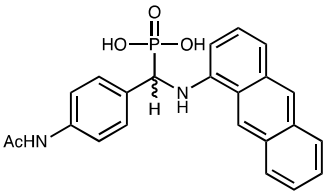
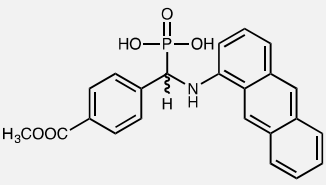
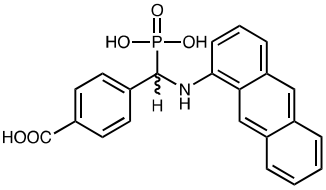


Figure 3.5. Non-linear curve fitting of compounds in R_1 screening

Table 3.3. List of α-aminophosphonate compounds and their IC_{50} against TcTS for R_1 screening

Compound	*Structure*	*LogIC50*	*Stand error LogIC50*	*IC_{50} (mM)*
49		-0.5743	0.053	0.27
50		-0.151	0.030	0.71
51		-0.536	0.021	0.29
52		0.371	0.132	2.35
53		-0.638	0.047	0.23
54		-0.677	0.084	0.21

Compound	*Structure*	*LogIC50*	*Stand error LogIC50*	*IC_{50} (mM)*
55		-0.376	0.102	0.42
56		-0.193	0.056	0.64
57		-0.585	0.098	0.26
58		-0.619	0.087	0.24
59		-1.816	0.027	0.015
60		-1.229	0.043	0.059
61		-1.663	0.03	0.022

Halogen bonding plays an increasingly important role in modern drug design because of its unique chemical properties (Lu *et al.*, 2012). However, the modification of halogens on R_1 groups did not show a positive effect for aminophosphonate compounds when inhibiting TcTS activity. Although the IC_{50} was improved for the 4'-triflouromethyl-phenyl- naphthalene compound (compound **51**), the inhibition effect become worse when trifluoromethyl group was added to different positions in anthracene compounds (compound **57, 58**). Chloride modification also resulted in weaker inhibition when added at the para position of the sialic acid moiety phenyl group (compound **50**).

3.4. Evaluation of specificity of α-aminophosphonate

3.4.1. Expression, purification, and characterization of bacterial sialidases

Purified *Clostridium perfringens* neuraminidase (CpNA) was purchased from sigma. The expression conditions for *Staphylococcus pneumoniae* neuraminidase A (SpNA) have been screened by the group of Prof. G. Taylor (University of St Andrews, UK) from whom the plasmid was obtained as a gift. 1 L expression in *E. coli* was performed by adding IPTG until concentration to 0.5mM then incubated at 25 °C for 14-16 hours. Following cell lysis, His-trap purification resulted in a single peak e.g. from fraction A13 to B2 (Figure 3.6 A), the fractions were then evaluated by SDS-PAGE (Figure 3.6 B). Compared to TcTS, SpNA shows a higher level of both target product (represented by the band at 74 KDa) and overall protein expression. Therefore, large numbers of impurity proteins were still eluted with target product, resulting in only about 60-65% purity. Secondary purification using anion-exchange chromatography effectively reduced the impurities. 17mg pure SpNA was purified from 1L culture.

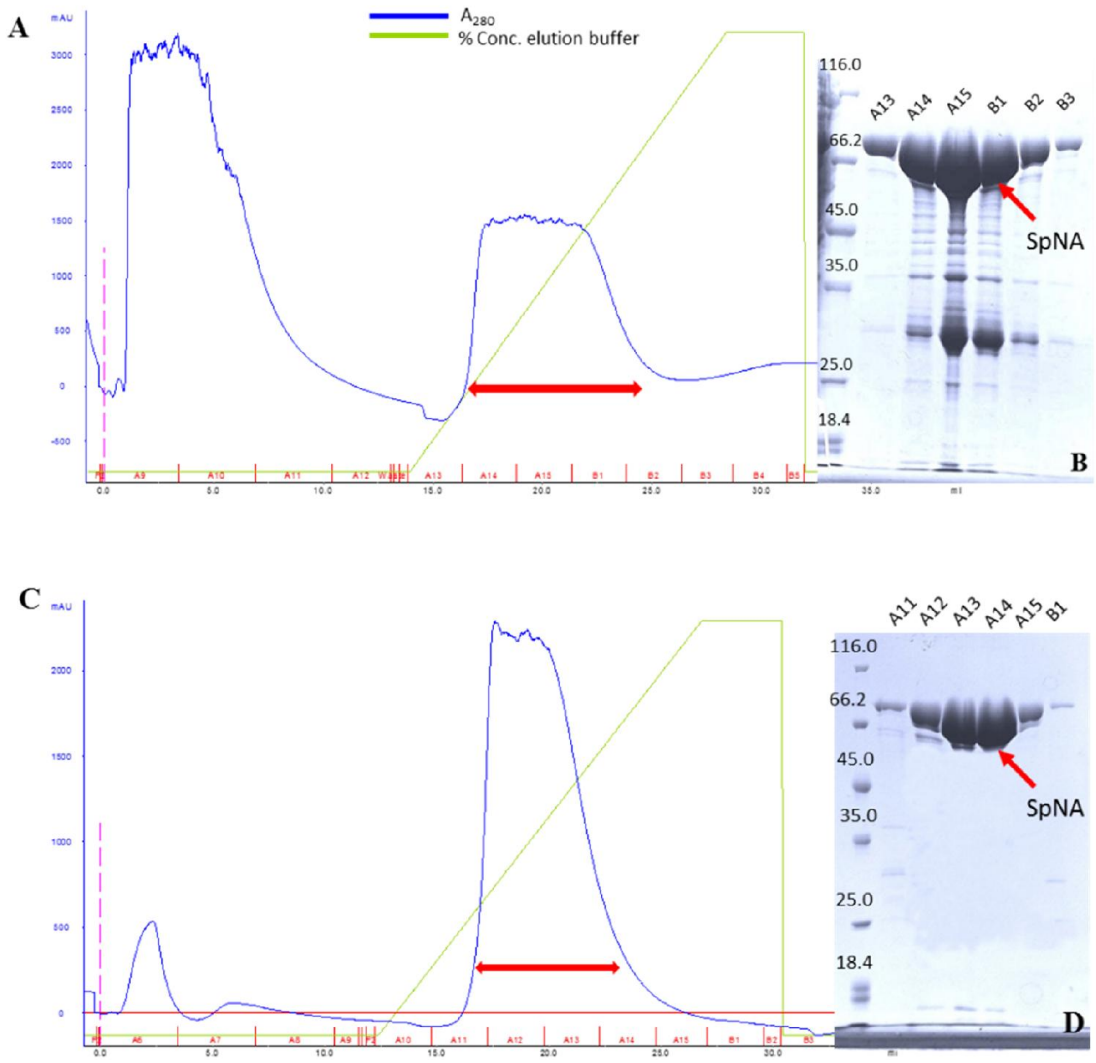


Figure 3.6. Chromatograph of His-trap purification of SpNA (**A**) and purity shown by SDS-page gel (**B**). Ion-exchange purification of SpNA (**C**) and SDS-PAGE gel (**D**).

CpNA has a K_m value of 0.15 mM binding with MuNANA, which is similar to TcTS (Chapter 3.2). SpNA shows a higher K_m to MuNANA (K_m=0.041mM). In terms of the turnover number, the k_{cat} value of CpNA cannot be determined because the concentration is unknown (sold by Unit, not concentration). The k_{cat} of SpNA is significantly higher than TcTS using MuNANA as substrate. However, as a non-natural substrate, using K_m of MuNANA to assess the K_m of natural substrates is still difficult, but these experiments give a general idea of the kinetic features of three different enzymes and is helpful for deciding the conditions of further inhibition assays.

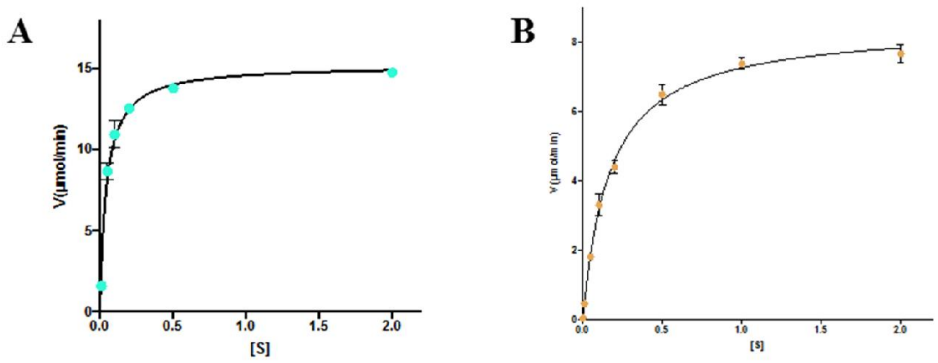


Figure 3.7. Non-linear curve fitting of rate against substrate concentration assay of SpNA (**A**) and CpNA (**B**)

Table 3.4. Kinetics parameters of SpNA and CpNA using MuNANA as substrate.

	K_m(mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})
SpNA	*0.041±0.002*	*109.3±2.5*	*2.66×10^3*
CpNA	0.15±0.02	N/A	N/A

3.4.2. α-aminophosphonate inhibit against bacterial sialidases

For a novel synthesized inhibitor, specificity is another important factor, particularly for non-natural molecules. In this section, the inhibition effect of α-aminophosphonate compounds was determined against two different bacterial sialidases SpNA and CpNA, both have similar structure to TcTS in the catalytic domain with a 6 bladed β-propeller, and an extra lectin-like domain. But the lectin-like domain for SpNA and CpNA are placed differently from TcTS relative to the catalytic domain and have carbohydrate binding activities (Yang *et al.*, 2015).

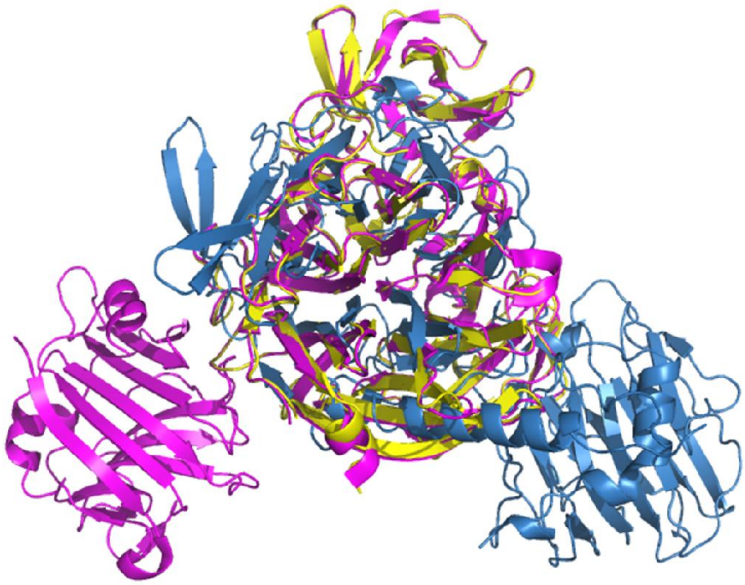


Figure 3.8. Structure alignment of the catalytic domains of TcTS (Blue, PDB: 1MS1), SpNA (pink, PDB: 2VVZ and 4C1X) and CpNA (Yellow, PDB: 2VK5) showing the different positioning of the lectin-like domains, generated by PyMol.

The IC50s of good α-aminophosphonate candidates (compound **48, 54, 59,** and **61**) were determined against SpNA and CpNA (Figure 3.9, 3.10 and Table 3.4, 3.5). IC_{50}s of SpNA and CpNA were determined with the same concentration of TcTS and same temperature, but different buffer was used to achieve the optimized activity of bacterial sialidases. The α-aminophosphonates show inhibition against bacterial sialidases with IC_{50}s in the high micromolar levels, which are all around 8-10 times higher than their IC_{50}s against TcTS (Figure 3.11). For example, compound 20 shows an IC_{50} of 0.015 mM against TcTS, and for the bacterial sialidases, it has 0.27 and 0.25 mM against CpNA and SpNA respectively. The results of bacterial sialidases inhibition indicates that the α-aminophosphonates are more specific inhibitors of TcTS rather than bacterial sialidases, meaning that the α-aminophosphonates could not inhibit other sialidases activity in the concentration at which they can inhibit TcTS. However, even if they act as a weak inhibitor for bacterial sialidases, they can still reduce the activity of these enzymes. Interestingly, for all three enzymes, the naphthalene compounds have much weaker inhibition effects than the anthracene compounds, meaning that the number of

aromatic cores is important for the inhibition.

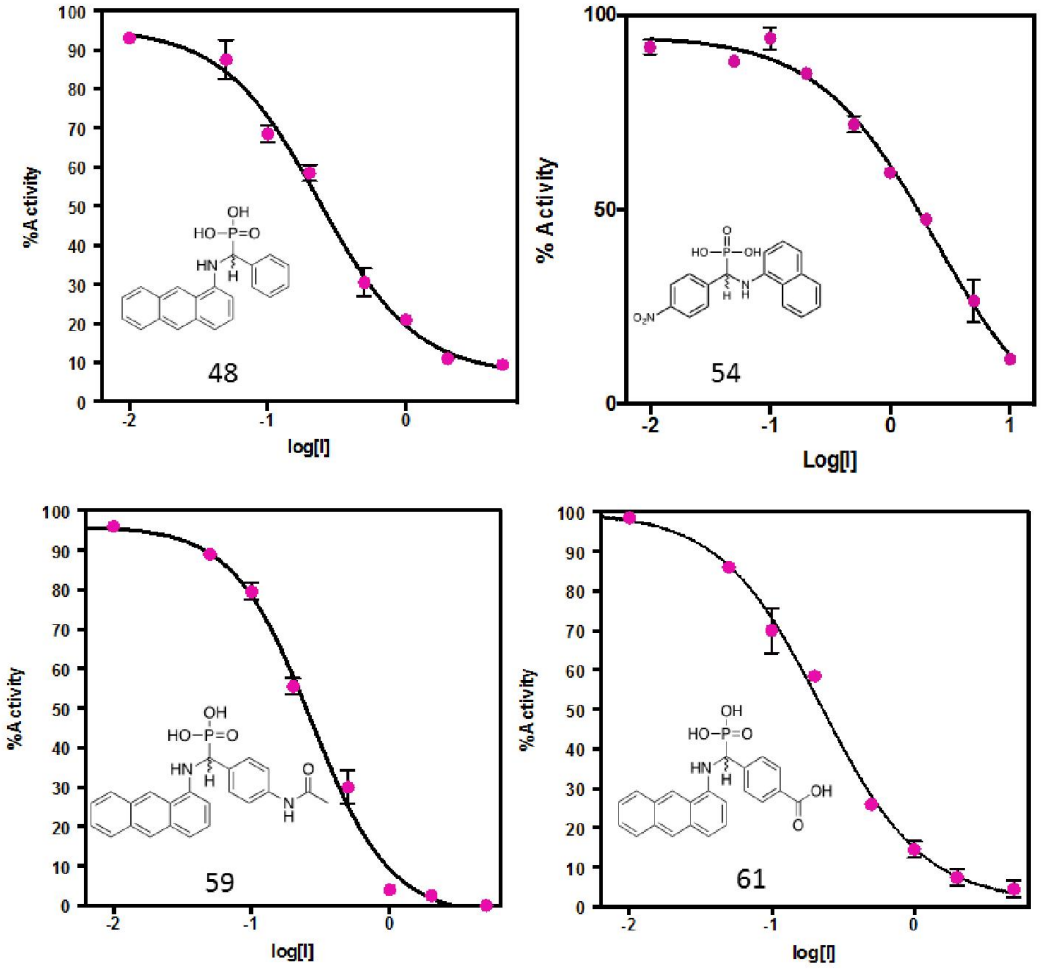


Figure 3.9. Non-linear curve fitting of IC_{50} of inhibitors against SpNA

Table 3.5. Summary of IC_{50} of inhibitors against SpNA

Compound	$LogIC50_{SpNA}$	Std. error $LogIC50_{SpNA}$	$IC50_{SpNA}$ (mM)
48	-0.625	0.041	0.237
54	0.378	0.086	2.388
59	-0.651	0.031	0.223
61	-0.557	0.028	0.271

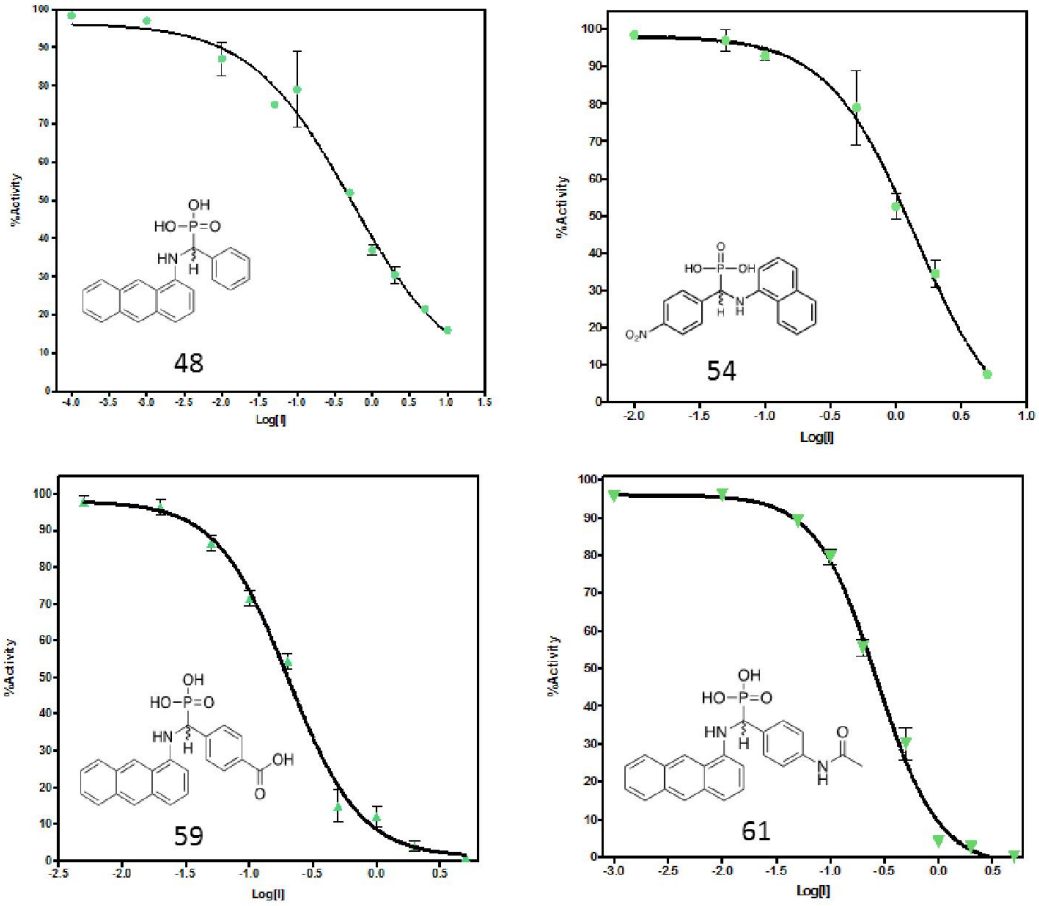


Figure 3.10. Non-linear curve fitting of IC_{50} of inhibitors against CpNA

Table 3.6. Summary of IC_{50} of inhibitors against CpNA

Compound	*$LogIC50_{CpNA}$*	*Std. error $LogIC50_{CpNA}$*	*$IC50_{CpNA}$ (mM)*
48	-0.29	0.11	0.51
54	0.15	0.052	1.41
59	-0.56	0.028	0.27
61	-0.66	0.033	0.22

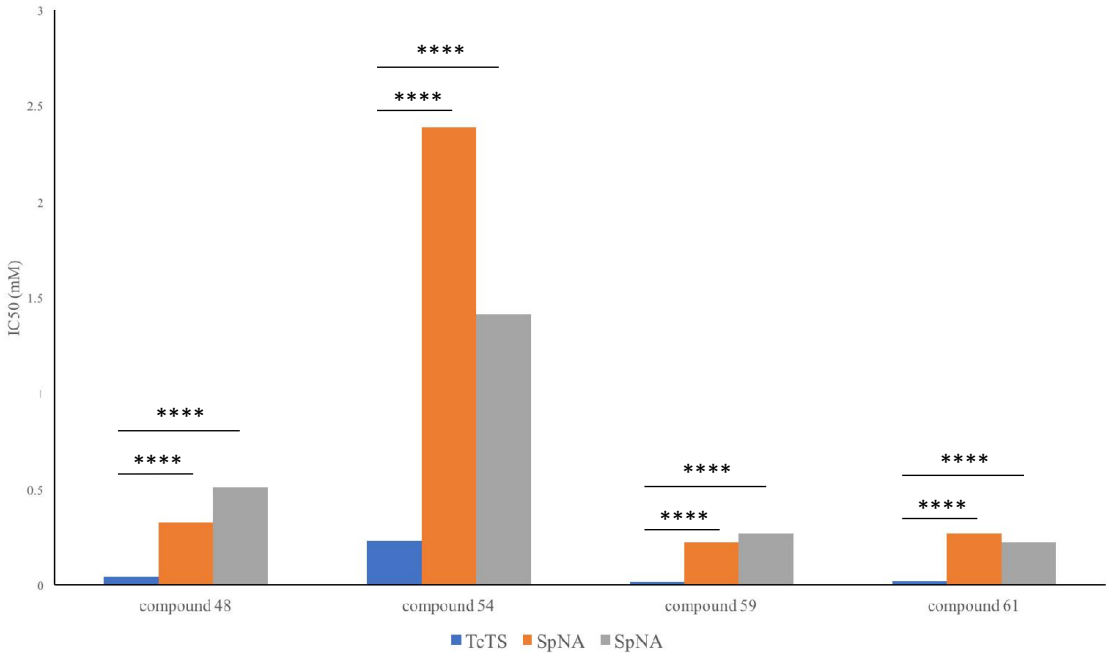


Figure 3.11. Compilation of IC_{50} results of α-aminophosphonate compounds against TcTS, SpNA and CpNA. Ordinary one-way ANOVA has been calculated by PRISM ($p<0.05$)

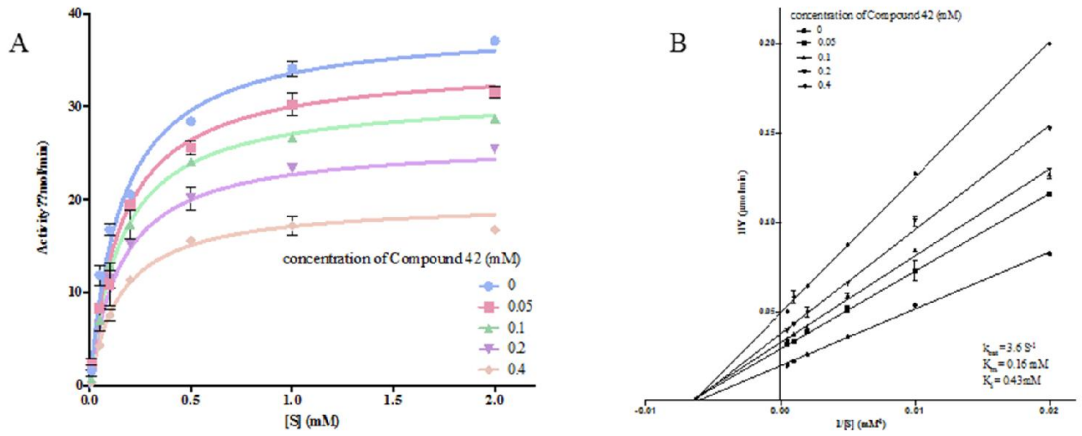
3.5. Mode of inhibition of α-aminophosphonate

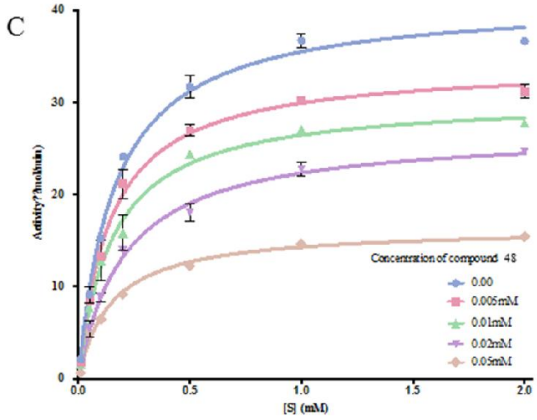
The inhibition assays show that α-aminophosphonate can be a potential specific TcTS inhibitor as it presents a higher inhibition effect against TcTS than other (bacterial) sialidases. Therefore, the mechanism of inhibition becomes important to understand the inhibition process and improve the inhibition effect. The mode of inhibition of these selected compounds was tested by enzyme kinetics using MuNANA as substrate. K_m and Vmax were determined with and without inhibitors, and the mode of inhibition identified from the change of K_m and V_{max} with increasing amount of inhibitors. The inhibition constant K_i can also be calculated using the appropriate inhibition model.

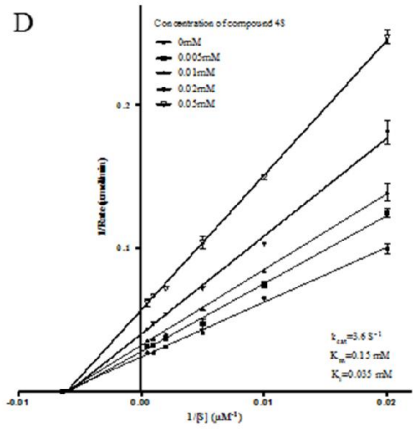
The K_m and Vmax of selected compounds were tested with differing concentration of inhibitors. Surprisingly, all detected compounds show a reduction of V_{max} with increasing inhibitor concentration, whilst the K_m values were remaining at the same level (Figure 3.12). The characteristics indicate that α-aminophosphonate is a non-competitive inhibitor of TcTS, interacting with TcTS by binding somewhere else than

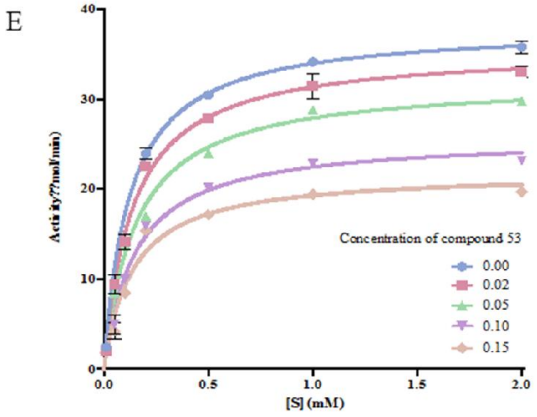
the catalytic pocket. The kinetic results were also fitted to other models such as competitive inhibition, mixed inhibition or uncompetitive inhibition, but all of them gave large standard error values, suggesting that non-competitive inhibition is the best fitting model for the inhibition process of α-aminophosphonates against TcTS.

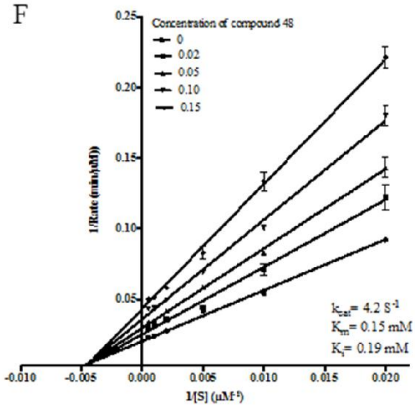
Since all the compounds investigated show non-competitive inhibition of TcTS, this suggest that the binding of this scaffold is not as predicted. Possibly, the simplification of the cyclical saccharide structure to an aromatic ring cannot be recognized by TcTS, or, unlike other sialidases which can recognize the phosphate group in the inhibitor molecules, the TcTS catalytic site can only recognize the carboxyl group at that position. Previous reports also identified pyridoxal 5'-phosphate as a weak and non-competitive inhibitor of TcTS. Interestingly, the α-aminophosphonates inhibit CpNA by a competitive mechanism, with the V_{max} unchanged but increasing K_m whilst increasing the concentration of inhibitor (Figure 3.13), and they also present mixed inhibition against SpNA. These kinetics results with bacterial sialidases show the ability of α-aminophosphonates to bind the sialidase catalytic site, which is also conserved in TcTS, but it is unable to be recognized as a substrate mimic in TcTS, suggesting a unique substrate recognition mechanism in TcTS.

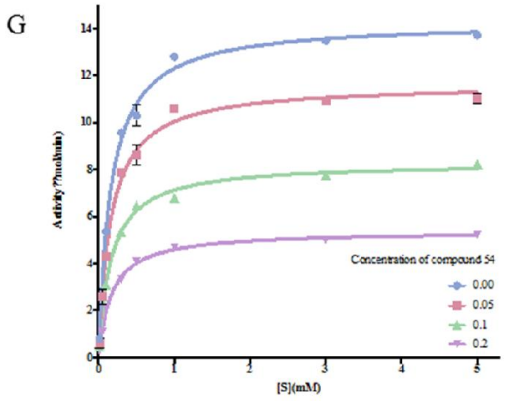


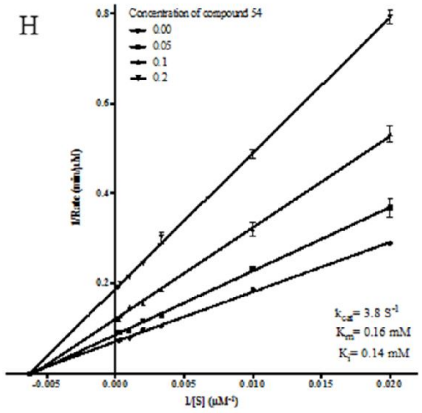

C
Concentration of compound 48
0.00
0.005mM
0.01mM
0.02mM
0.05mM
[S] (mM)

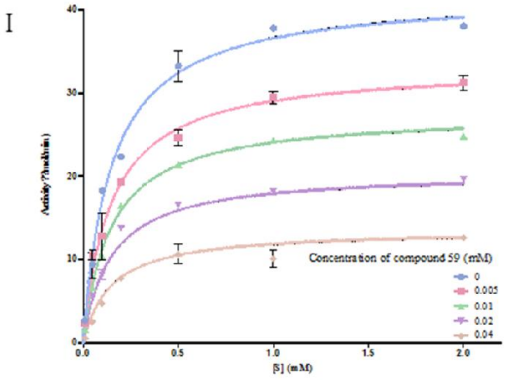

D
Concentration of compound 48
0mM
0.005mM
0.01mM
0.02mM
0.05mM
1/[S] (μM-1)

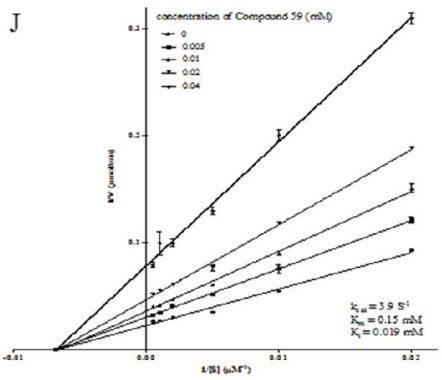

E
Concentration of compound 53
0.00
0.02
0.05
0.10
0.15
[S] (mM)


F
Concentration of compound 48
0
0.02
0.05
0.10
0.15
1/[S] (μM-1)


G
Concentration of compound 54
0.00
0.05
0.1
0.2
[S](mM)


H
Concentration of compound 54
0.00
0.05
0.1
0.2
1/[S] (μM-1)


I
Concentration of compound 59 (mM)
0
0.005
0.01
0.02
0.04
[S] (mM)


J
concentration of Compound 59 (mM)
0
0.005
0.01
0.02
0.04

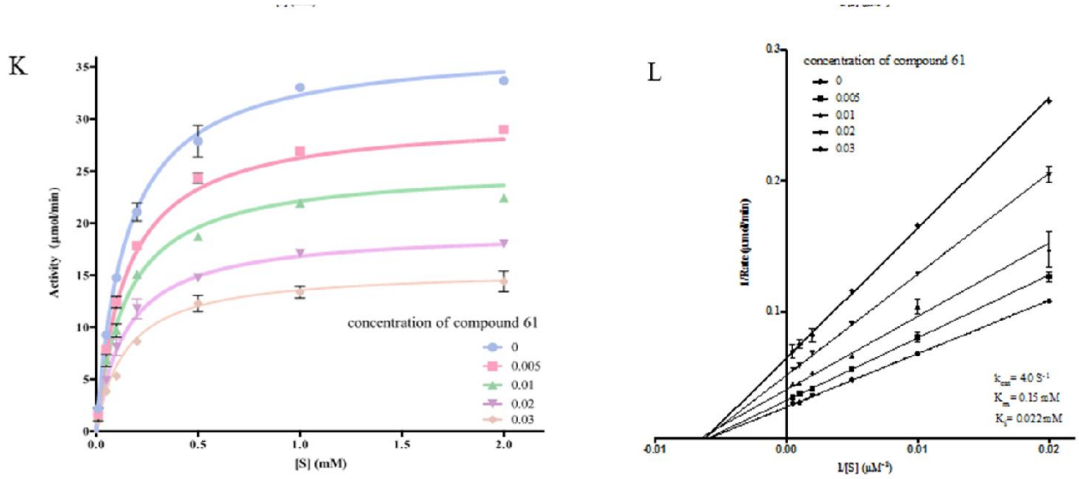


Figure 3.12. Non-linear fit (left) and Lineweaver–Burk plot (right) of TcTS activity with different concentration of α-aminophosphonate compound **42 (A,B), 48(C, D), 53 (E, F), 54(G, H), 59 (I, J), 61 (K, L)**.

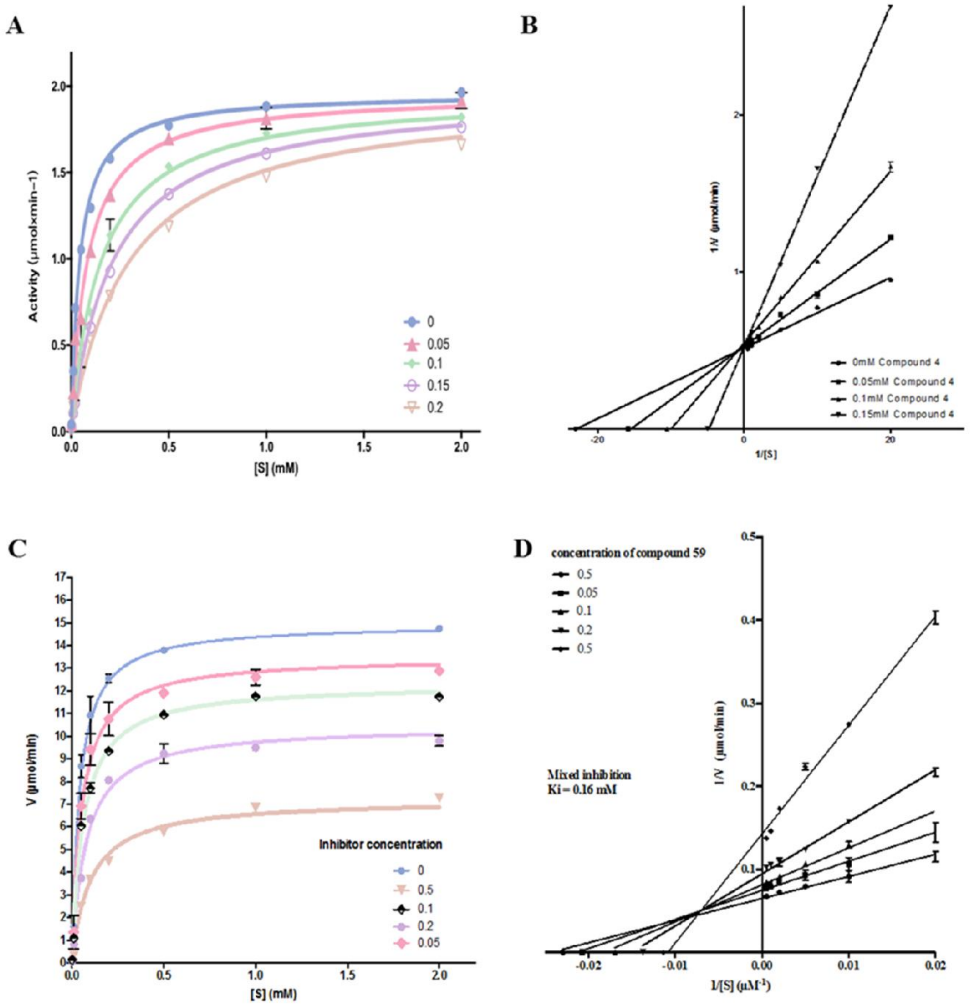


Figure 3.13. Non-linear fit and Lineweaver–Burk plot of CpNA (**A, B**) and SpNA (**C, D**) activity with different concentrations of α-aminophosphonate compound **59**.

3.6. Time-dependent inhibition of α-aminophosphonate against TcTS

For most reversible inhibition including competitive, non-competitive and mixed inhibition, the inhibition effect does not normally change with increasing time of incubation *in vitro* or dose *in vivo*. While many irreversible inhibitions, for example, covalent inhibition and aggregation-based inhibition, may present time-dependent inhibition characteristics (Riley *et al.*, 2007). Therefore, the time dependency of the inhibition effect can be an important factor in drug discovery and development to evaluate the authenticity of inhibition, to avoid the fake positive of non-specific aggregation-based inhibition (Feng and Shoichet, 2006).

The time dependent inhibition (TDI) assay was based on the IC_{50} assay but the time of pre-incubation of TcTS with inhibitors was changed. As in the results shown in Figure 3.11, the IC_{50}s of two naphthalene compounds (compound **41** and **52**) and one 1-anthracene compound (compound **48**) were determined with pre-incubation times of 0, 10, 30, 60 and 300 mins. Generally, the demand of inhibitor concentration to inhibit the activity of TcTS were kept in same level with the pre-incubation time up to 5 hours.

In general, α-aminophosphonate compounds do not show time-dependent inhibition against TcTS, which means the inhibition is not affected by the incubation time. Although the IC_{50}s without pre-incubation of the three inhibitors were higher than samples with longer pre-incubation, this might probably be because of the lack of sufficient time for inhibitors to interact with TcTS prior to being challenged with substrate. Therefore, from the results of these TDI studies, rather than an irreversible or covalent inhibitor, α-aminophosphonate compounds seem to be reversible inhibitors of TcTS.

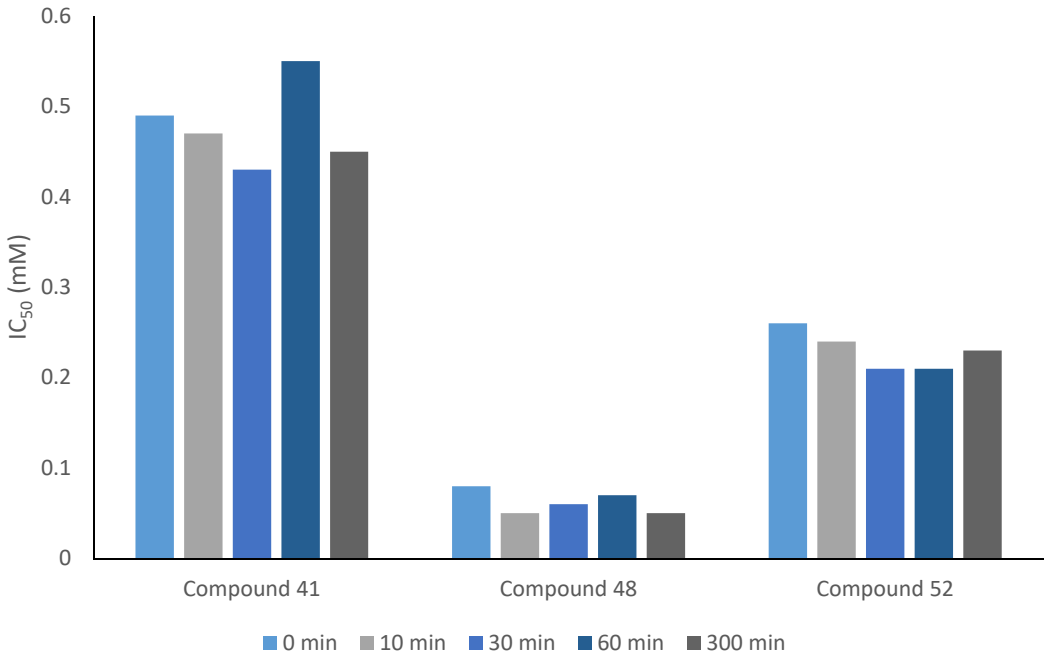


Figure 3.14. IC_{50}s of compound **41, 48, 52,** with different incubation time

Table 3.7. IC_{50}s of compound **41, 48, 52** with pre-incubation from 0 min to 5 hours.

Compounds	*0 min*	*10 min*	*30 min*	*60 min*	*300 min*
	IC_{50} (mM)				
Compound 41	0.51	0.47	0.43	0.55	0.41
Compound 48	0.08	0.05	0.06	0.07	0.05
Compound 52	0.26	0.24	0.21	0.21	0.23

3.7. Thermodynamic study of α-aminophosphonate-TcTS binding

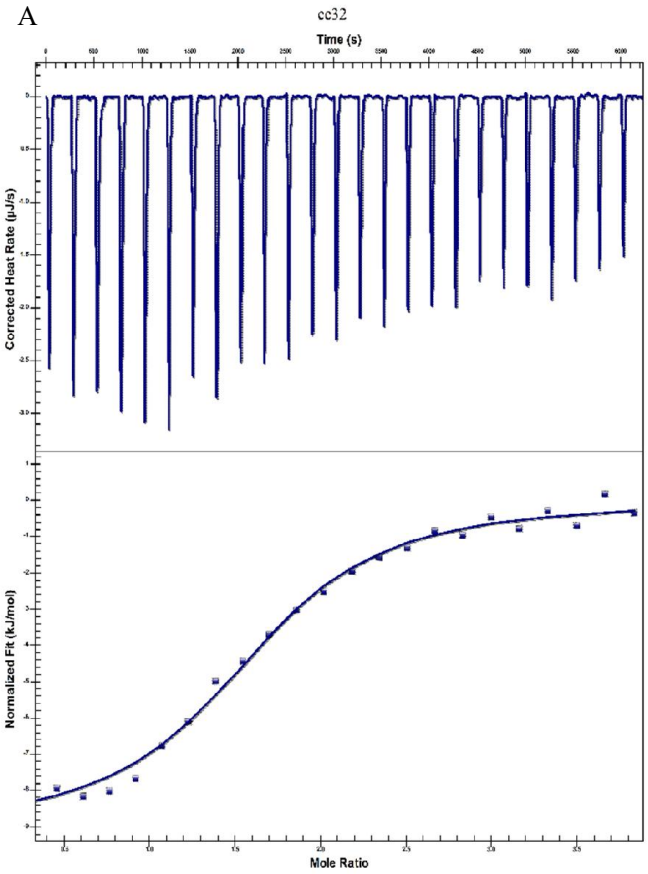
Thermodynamic characterization, mainly by isothermal titration calorimetry (ITC) is useful in drug candidate evaluation, and both binding constants and energy changes during the interaction may be obtained. However, some limitations result in not all protein-ligand or protein-protein interactions being detectable by ITC.

In an ITC experiment, a large amount of protein sample at high concentration (10-50 times the K_d) is required to achieve a full titration curve (Damian, 2013). The demands of high protein concentration and K_d ratio mean that only high affinity interactions can be determined accurately by ITC. Due to these limitations, although several promising TcTS inhibitor candidate have been developed, none of them have published thermodynamic data, since the best inhibitor against TcTS only have IC_{50} value in micromolar level, which is not equal but close to K_d. However, an expected ITC experiment require 50-10 of Protein concentration: K_d ratio (Duff *et al.*, 2011), results in which the protein concentration is not easy to reach the ITC experimental requirements.

Pure TcTS is only stable at concentrations up to 12 mg/ml (168.3 μM) in either Tris buffer or PBS, and it tends to aggregate in solution during ITC experiments (data not shown). ITC experiment was performed using NanoITC (TA instrument). Purified and degassed TcTS was injected into the sample cell, the ligand solution was loaded by syringe and titrated into the sample cell with 250rpm stirring. The temperature change for each titration was recorded and a 270 second gap was set between two titrations. The results were analysed by NanoAnalysis software (TA instrument) and thermal parameters of the interaction were calculated. Several α-aminophosphonate derivatives were tested in ITC assays but with most of them the full titration curve was not achieved due to the high K_d. For compound **59**, the compound with best inhibition effect (IC_{50} =

15 μM) and high specificity, the saturation curve was successfully obtained in an ITC assay (Figure 3.16).

A



B

n	1.359±0.25
Kd (μM)	12.09±4.46
ΔH(kJ/mol)	-10.52±0.05
ΔS(J/mol·K)	58.37
-TΔS (kJ/mol)	-17.37
ΔG	27.89

C

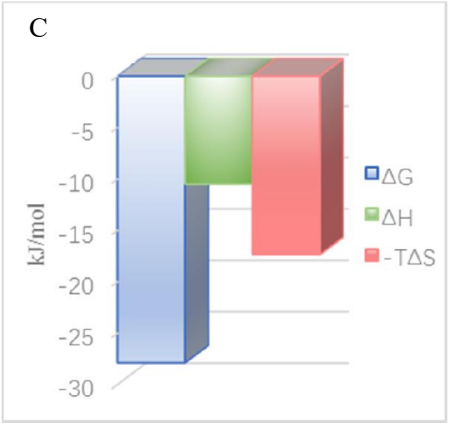


Figure 3.15. A. Experimental data (up) and non-linear curve fitting (down) of ITC experiment, in which 5mM of compound 59 was titrated into 0.16 mM of purified TcTS. **B.** Thermodynamic parameters calculated from ITC experiment. **C.** A bar chart to present ΔG, ΔH, and –TΔS.

The ITC result gave a binding constant K_d of compound **59** at 12 μM, which support the inhibition parameters (IC_{50} and K_i) obtained from kinetic assays, and suggest that the interaction between TcTS and compound **59** is a entropy driven spontaneous process, since entropy has higher contribution to the total binding energy than enthalpy. The ITC result also indicates a 1:1 binding ratio by the n value between TcTS and inhibitor. However, the interacting position and inhibition mechanisms are still unclear since the binding site of the α-aminophosphonate is still unknown.

3.8. Discussion

3.8.1. Molecular docking is unsuccessful in its prediction of TcTS-inhibitor interaction

The design of α-aminophosphonate compounds was based on the structure of TcTS substrates. However, it has been shown experimentally that they all show non-competitive inhibition mechanisms, meaning that this class of compounds do not interact with TcTS at the catalytic pocket. Although molecular docking simulations shows that this class of compounds have high affinity to the catalytic site of TcTS, kinetics and thermodynamics results suggest that, in this case, the docking results are not reliable.

Comparing the results of molecular docking and enzyme kinetics assays, it can be seen that molecular docking might not be a suitable technique to predict or investigate the binding of ligands when the binding site or mode of binding is unknown, and the affinity predicted by molecular docking cannot be as accurate as the experimental data. To increase accuracy of a rigid docking simulation, molecular docking is performed in a limited area, In the case of TcTS, no secondary binding sites, only the catalytic site, have been reported, hence the difficulty of predicting possible binding sites which have not been proved through experiments. More importantly, the macromolecule model in the molecular binding protocol used cannot perform major conformational changes, even conformational changes of Cα chains. For most docking programs, the flexibility of residues is limited to the side chain. Thus, interactions with ligands which may lead to conformational changes will be difficult to predict by molecular docking. Additionally, molecular docking does not consider the solvent environment of the macromolecule, this can also reduce the accuracy of prediction.

However, even though molecular docking failed to predict the interaction between

TcTS and α-aminophosphonate compounds, it is still a useful tool in drug development. For cases in which the binding site of drug candidate is known, and the interaction does not lead to major conformational changes, such as most competitive inhibitors, molecular docking can be used to interpret the functions of different groups in drug candidates, to then improve their effect and specificity by rational modifications. Polar or charged substitutions on the 4' position of the SA-mimic phenyl give a significant increase in inhibition and also enhance the solubility of the compound. However, only a phenyl group has been screened as the SA-mimic moiety in this study, multiple aromatic rings structure such as naphthalene or anthracene could be tested.

3.8.2. Influence of R_1 and R_2 α-phosphonate substitutions on the TcTS inhibition effect

Although now known to be a non-competitive inhibitor, different substitutions in both R_1 and R_2 do still affect the inhibition, by some other mechanism than the initial design. Therefore, it is more suitable to describe the two moieties of the α-aminophosphonate compounds as the aldehyde (AL) moiety and the amine (AM) moiety, which are based on the reagent in synthesis, rather than "sialic acid moiety" and "lactose moiety" respectively. Although the kinetic results suggest a different mode of inhibition from original design, the screening results of both the sialic AL moiety and AM moiety still a structure-activity relationship that fitted the original model, for example, an anthracene at the amine moiety shows stronger inhibition than naphthalene compounds. Although the actual binding site of α-aminophosphonate compounds is not the catalytic site, this suggests they might have similar interactions.

For the AM moiety, the screening in this study mainly focused on the type of aromatic group but not the modifications on these aromatic groups. Some characteristics can be drawn from the AM moiety screening. Firstly, direct linkage of aryl groups at the amino nitrogen gives stronger inhibition then linkage with extra carbon (compare between compound **42** and **47**), that may be induced by increased flexibility result from the extra

chiral center. And the increasing number of aromatic rings enhance the inhibition effect. In terms of the AL moiety, adequate modification of the benzene ring on AL group can be helpful for inhibition, for example, the introduction of nitro, carboxyl and acetamido group enhance the inhibition effect. However, the substitutes containing halogens in AL moiety gave limited improvement or even negative effect for the inhibition.

3.8.3. Hypotheses of inhibition mechanism

As a non-competitive inhibitor, the K_m value between substrate and TcTS does not change when the α-aminophosphonate is bound. Therefore, the location of the allosteric binding site and the mechanism of how the α-aminophosphonates inhibit the activity of TcTS become important in further investigation and drug development. Two hypotheses come from the inhibition study and TcTS structure characteristics. Firstly, the binding of the inhibitor could stop hydrolysis progress by effecting key residues in the catalytic pocket. Non-competitive inhibition generally means that the K_m value of substrate is not affected by inhibitor binding, but the catalytic process has been impacted (Su *et al.*, 2008). Except the catalytic site, no other binding sites on TcTS for substrate or inhibitor has been reported. It can be expected that α-aminophosphonate derivatives will not interact with residues which directly bind to substrates such as the arginine triad, Tyr342, Trp120, and Asp59, but it could limit the catalytic process by effecting other key residues, for example, locking Tyr119 in either the closed and open position.

Secondly, the inhibition could be achieved by changing the rigidity of the catalytic site. The catalytic site, especially the lactose binding site of TcTS changes conformation to release the cleaved donor polysaccharide chains and accept the acceptor polysaccharide moiety, to achieve the transfer activity and avoid self-adhension.

3.8.4. Conclusion

In short, α-aminophosphonate compounds inhibit TcTS, but show an unexpectedly non-competitive manner and inhibit TcTS activity by influence of catalytic process rather than effect donor substrate binding, since it does not change the K_m of MuNANA. It can be hypothesized that interaction between α-aminophosphonate and TcTS affect the conformation or rigidity of the catalytic site, result in eliminating either hydrolytic activity or catalytic product release. A structural study of TcTS-inhibitor complex would be helpful to interpret the position of the allosteric binding site and inhibitory mechanisms

Chapter 4. X-Ray crystallography study of TcTS in complex with α-aminophosphonate inhibitors

4.1. Prelude

Based on the published crystal structures of TcTS binding with its substrate sialyl-lactose, α-aminophosphonates were designed and screened as selective TcTS inhibitors. However, through examining the mode of inhibition, all tested α-aminophosphonate candidates were shown to inhibit in a non-competitive manner, indicating that the inhibition of α-aminophosphonate follows an allosteric process. By biophysical and biochemical investigations alone, it is difficult to define the binding site and detailed inhibition mechanisms. Therefore, structural studies of TcTS in complex with inhibitors are important and necessary for investigation of inhibitory mechanisms and further drug development.

This chapter describes structural studies of TcTS-inhibitor complexes, including α-aminophosphonate compounds and published inhibitors such as DANA, lactitol and Rhein. Since the α-aminophosphonate scaffold does not inhibit TcTS activity through mechanism-based inhibition, it is crucial to understand the binding position of this new type of inhibitor and investigate the inhibition mechanism based on the allosteric binding. Hitherto, no structure of non-competitive inhibitors in complex with TcTS has been reported and the detailed inhibition mechanisms of these compounds are still unclear. Our structure of a TcTS-aminophosphonate complex is the first reported structure of TcTS in complex with a non-competitive inhibitor, and a new binding site between the catalytic domain and lectin-like domain was identified by the crystal structure, which is critical in the investigation of TcTS mechanism and development of further improved inhibitors.

The inter-domain allosteric binding site is formed by the movement of several residues, which then form a generally positively-charged binding pocket for the aminophosphonate inhibitors. To validate this novel binding site, two compounds which had an anthraquinone group replacing the anthracene were introduced, although

with reduced inhibition effect, they bind to the same binding site as the anthracene compounds. Meanwhile, an anthraquinone fragment was also defined in the lactose-binding site, and this binding was also validated by the crystal structure of TcTS soaked with Rhein, another known non-competitive anthraquinone inhibitor of TcTS.

Further mutagenesis of residues, which interact with inhibitors in crystal structure, in the allosteric site was also examined. By truncating the side chains of three key arginines, the inhibition was lowered, but the overall activity was not greatly affected although the K_m and k_{cat} were slightly changed. Through bioinformatics studies of the catalytic site and allosteric site of TcTS, other trans-sialidase and sialidases were also studied to explore why α-aminophosphonate is a selective TcTS inhibitor.

4.2. Auto induction procedure to increase the yield of TcTS expression

Initially, expression and purification of recombinant TcTS was performed as described in **Chapter 3**, using IPTG induction and a two-step purification including His-trap and anion exchange chromatography. The purity after ion-exchange chromatography was adequate for high-purity assays, such as protein crystallography and ITC. However, the yield following IPTG induction was normally 3-4 mg pure TcTS per liter of *E. coli* culture. For crystallography, which requires larger amount of high-purity protein, IPTG induction alone was insufficient.

Therefore, to increase the yield of protein purification, firstly TB media was tried as a replacement for LB media, an enrich media with additional nutrients and growth factors which potentially increase yield of *E. coli* (Tartof and Hobbs, 1987), but a similar yield of protein was obtained. Additionally, autoinduction was introduced as an alternative method to IPTG induction. After screening of different media, temperature and length of incubation time, autoinduction for 48 hour in TB media at 16°C was shown to give

the highest yield of 18-21 mg of pure protein per liter of *E. coli* culture after two-step purification.

4.3. Crystallization of TcTS

TcTS crystallization trials has been set either using robotic screening (Phoenix nano dispenser) of various conditions (JCSG, PEG screening from Molecular Dimensions), or manual screening based on the published conditions. Micro crystals grew from **condition A** (100mM Tris-HCl, 20% PEG 4000, 5% isopropanol, pH 7.5) after a few days (Figure 4.1 A). However, it was difficult to harvest single crystals and the microcrystals were not big enough for soaking and data collection. While, a few single crystals grew in **condition B** (100mM HEPES, 20% PEG 4000, 5% isopropanol, pH 7.5), after few months (Figure 4.1 B), they still had the same problem with size since small crystals tend to dissolve when soaked with ligands which impacts the diffraction quality. However, the microcrystals were useful as seeds in **condition B**. After seeding, large single crystals, which are longer than 10nm and survived long soaking times were observed in the seeding drops (Figure 4.1 C, D, E) in a few weeks.

Co-crystallization conditions were optimized using 24-well plates. Single crystals were grew after a few weeks （Figure 4.2 A, B. These crystals were flash-cooled and sent to the Diamond synchrotron (beam-line I04) where data were collected to 0.9Å (Figure 4.2 C). Indexing of the data revealed a small P1 unit cell (unit cell dimensions: a = 9.90 Å, b = 10.09 Å, c = 16.41 Å, $\alpha = 78.56°$, $\beta = 79.35°$, $\gamma = 79.62°$) which would be very small for a protein, indeed the size of the unit cell is about the size of the compound alone. Analysis by ShelxC (Sheldrick, 2008) showed that the weight of the cell is 365Da, which is approximately the same as the ligand molecule (m.w. = 363 g/mol), however, the anomalous signal was too weak to solve the ligand crystal structure.

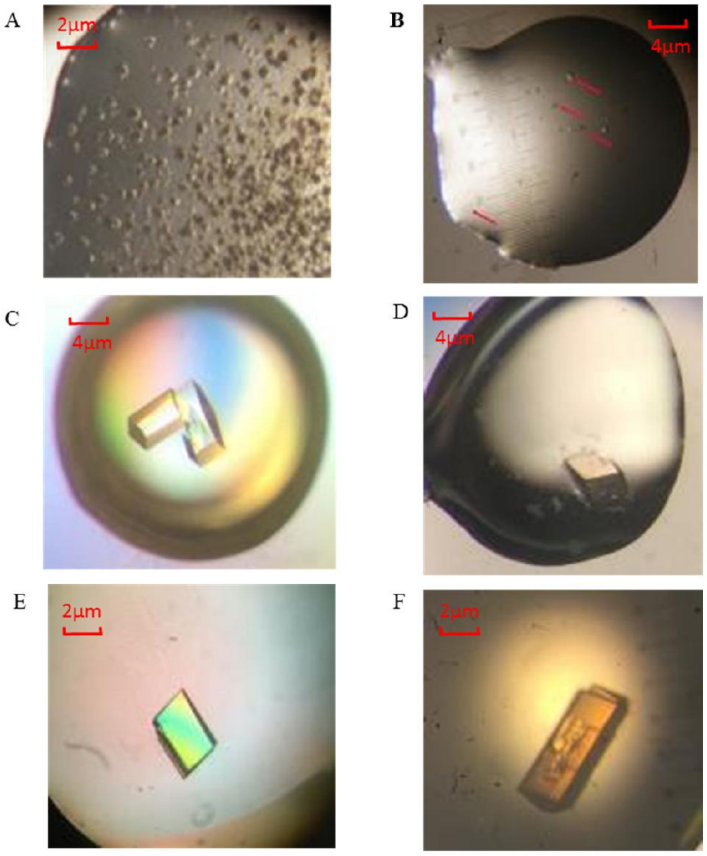


Figure 4.1. Gallery of TcTS crystals: **A.** Microcrystals, **B.** Single crystal grown in condition highlighted by red arrows, **C.** Single crystal grown by macro seeding in condition A; **D-E.** Single crystal using macro seeding in condition B. **F:** Single crystal soaks in **compound 59** for 48 hours

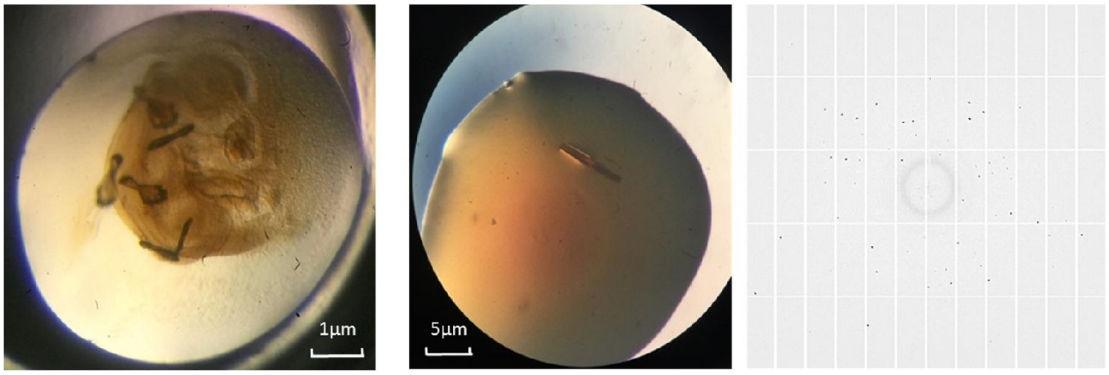
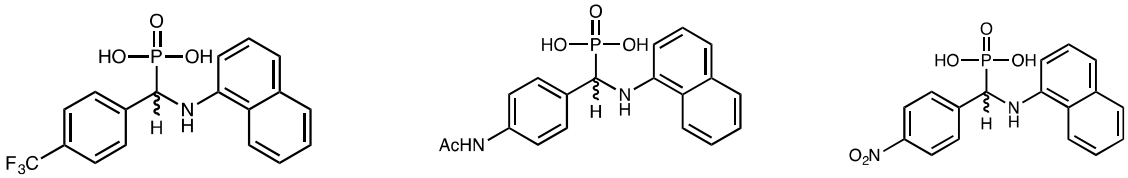


Figure 4.2. (**A**): soaking crystal into compound **59**. (**B**): A single crystal grew from co-crystallization of TcTS with compound **59**, in optimized condition. (**C**): Diffraction pattern of co-crystal of TcTS complex with compound **59**

4.4. TcTS crystal structure in complex with naphthalene inhibitors

In SAR screening, α-aminophosphonate compounds containing naphthalene gave a range of IC_{50} values from high micromolar to low millimolar levels, the compounds with 4'-nitrophenyl, 4'-acetamidophenyl and 4'-triflurомethylphenyl groups have the best IC_{50} of about 200μM, which are comparable to lactitol, the best commercially available TcTS inhibitor (Agusti *et al.*, 2004). Although naphthalene compounds were not the best inhibitors in screening, it is still interesting to investigate the how the structure of TcTS interacts with this series of compounds to further understand the structure-mechanism relationship of α-aminophosphonate inhibitors.



51 **53** **54**

Figure 4.3 Structure of naphthalene compounds that soaked with TcTS crystals

Single crystals of TcTS were soaked in crystallization solution (**Condition B**) containing TcTS and up to 10 mM of one of the compounds **51, 53 and 54** (Figure 4.3). A study of diffraction data from crystals soaked for different lengths of time by a project student Scott Gardner showed that positive difference peaks for soaked ligands started to appear in the electron density maps after 10 minutes soaking and crystals survived 48 hours soaking (Gardner, MSc thesis, 2014), therefore the length of soaking in subsequent studies were generally 48 hours. Details of data collection (in-house and Diamond synchrotron source) and refinement are shown in table 4.1. In monoclinic crystals in complex with **51**, positive peaks could be defined on the surface of the catalytic domain, at a distance of about 35 Å from the catalytic site. However, only the 4'-triflurомethylphenyl moieties of **compound 51** interacts with the catalytic domain

surface. The naphthalene moiety bound to the lectin-like domain of a symmetry-related molecule (SRM) (Figure 4.4), but without significant interactions with any residues in the lectin like domain, only one hydrogen bound with the main chain of Lys543 and two hydrophobic interaction with the C-β of were found with the (Figure 4.5). The same binding position was also seen in the monoclinic crystal of TcTS soaked with compound **54**. The binding regions on both the catalytic and lectin-like domains are flexible loops, but the interaction between TcTS and these compounds do not change the conformation of these loops greatly when compared to a structure without inhibitors and it is at some distance from the active site. Therefore, in these structures, it is difficult to judge how this interaction affects the activity of TcTS.

Table 4.1. X-ray data collection and model refinement statistics of TcTS crystal soaked with naphthalene compounds

Crystal No.	1	2	3	4
Beam line	In-house	Diamond beamline I03	In-house	Diamond beamline I03
Protein	TcTS	TcTS	TcTS	TcTS
Ligand soaked	**51**	**54**	**54**	**53**
Soaking time (h)	48	48	48	72
Ligand identified	**51**	**54**	**54**	N/A
Data collection and processing statistics				
Wavelength (Å)	1.541	0.976	1.541	0.976
Resolution (Å)	68.3-1.68	5.86-1.51	72.0-1.75	51.94-1.37
Space Group	P 1 2_1 1	P 1 2_1 1	P 2_1 2_1 2_1	P 1 2_1 1
Unit cell dimensions				
a (Å)	54.09	54.29	60.90	54.5
b (Å)	129.71	129.50	87.79	130.03
c (Å)	59.32	54.44	120.56	54.62
α (°)	90	90	90	90
β (°)	108.5	107.87	90	107.63
γ (°)	90	90	90	90
No. observations	81996	108821	67854	161507
Completeness (%)	89.9 (82.2)	95.41 (95.3)	92.1 (80.1)	99.8 (99.7)
Ave. redundancy	4.68 (3.24)	3.29 (2.8)	4.12 (3.10)	6.4 (5.3)
R_{merge}[a]	0.092 (0.482)	0.046 (0.841)	0.085(0.564)	0.052 (1.315)
I/σI	9.2 (2.4)	10.0 (1.2)	8.1 (1.3)	15.4 (1.2)

Refinement				
Reflections used	60306	98365	62482	82653
Number of protein residues	624	624	621	623
Number of water molecules	646	646	642	620
Number of atoms	10445	10189	9972	10224
R_{factor} [b]	0.17	0.12	0.21	0.19
R_{free} [c]	0.22	0.14	0.25	0.19
RMSD bond angle (°)	0.01	0.01	0.004	0.014
RMSD bond length	1.071	1.181	0.707	1.457

[a] $R_{merge} = \sum_{hkl}\sum_{j=1}^{N}\left|I_{hkl} - I_{hkl}(j)\right| / \sum_{hkl}\sum_{j=1}^{N}\left|I_{hkl}(j)\right|$

[b] $R_{factor} = \sum_{hld}\left|\left|F_{obs}(hkl)\right| - \left|F_{cal}(hkl)\right|\right| / \sum_{hld}\left|F_{obs}(hkl)\right|$

[c] R_{free} was calculated as defined by Brunger for 5% of the total data(Brunger, 1992)

*The calculation of R_{merge}, R_{factor} and R_{free} in following tables of data collection and refinement will follow the same method.

Interestingly, one of the crystals soaked with **compound 54** crystallized in an orthorhombic space group (P 2_1 2_1 2_1) which has a different crystal packing from the monoclinic crystals. In this crystal form, the catalytic site was blocked by a symmetry related molecule, therefore, it not suitable for study of ligand binding especially for ligands that might bind to the catalytic site. However, by comparing the binding site of different crystal forms, it is possible to determine whether the interaction between ligand and protein is real or merely a structure bias resulting from the crystal packing. In the orthorhombic crystal structure of TcTS soaked with **compound 54,** electron density, which can be identified as the inhibitor molecule, can still be found between adjacent symmetry related molecules but in a different place to that seen in the monoclinic crystal structure. In the orthorhombic crystal structure, model of **compound 54** cannot be identified on the interface between the catalytic domain and the SRM lectin-like domain, but on the interface between lectin-like domain of two SRMs (Figure 4.4 **E, F**). Crystal structures in different crystal form suggest that the binding

of naphthalene compounds on the interface between two SRMs are formed by crystal contact but not a real functional interaction.

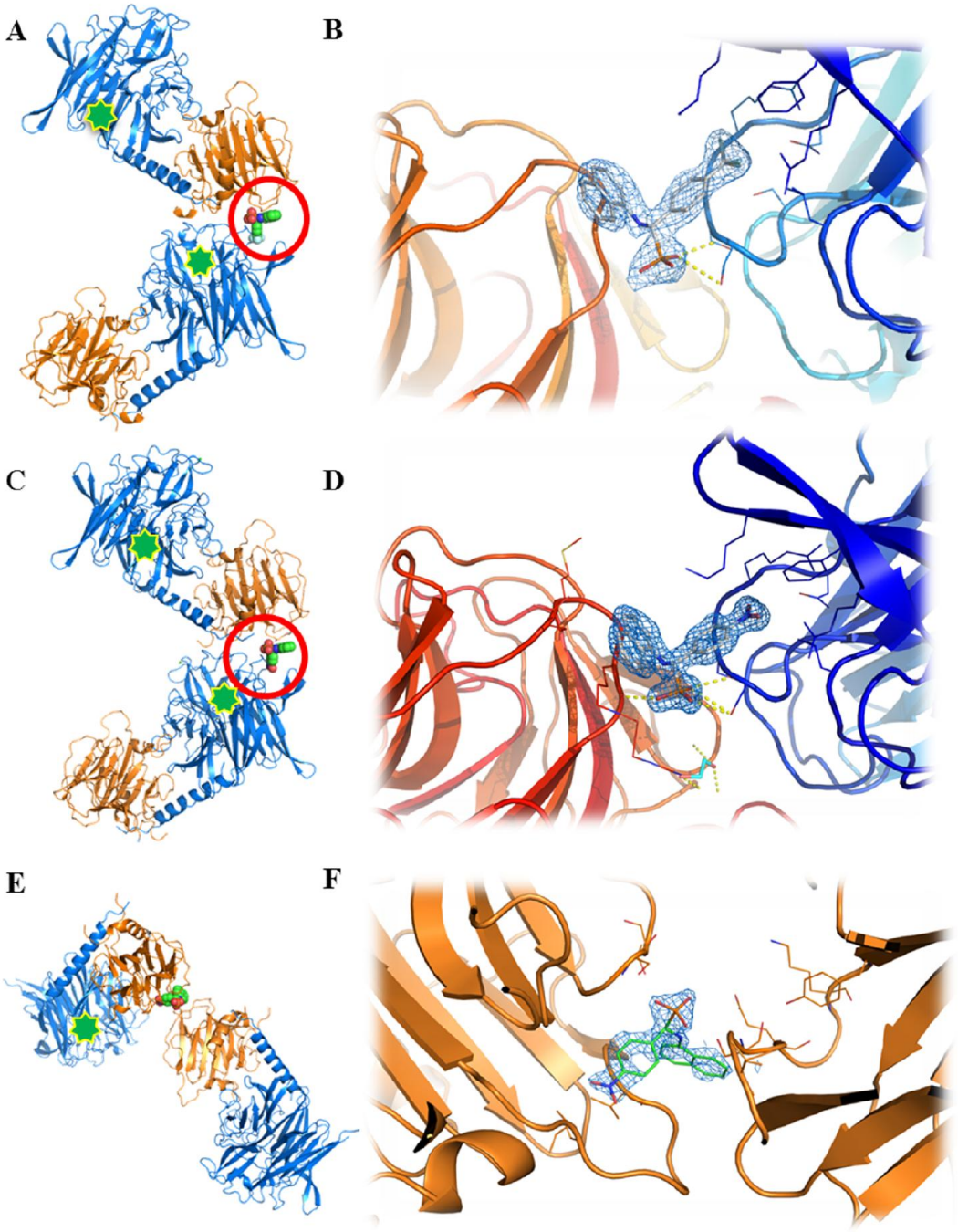


Figure 4.4. Inter-symmetry molecule binding position of **compound 51** in monoclinic crystal (P1 2_1 1) crystal **(A)** and detailed binding site with *2mFo-DFc* map (σ=1.0) **(B)**. Inter-symmetry binding position of **compound 54** in monoclinic crystal (P1 2_1 1) crystal **(C)** and detailed binding site with *2mFo-DFc* map (σ=1.0) **(D)**. **Compound 54** soaked into a P 2_1 2_1 2_1 crystal between two symmetry related molecule, binding position of compound 54 is highlighted by red circle and catalytic is labeled as green star (**E**). Binding detail of compound 54 with *2mFo-DFc* map (σ=1.0) (**F**).

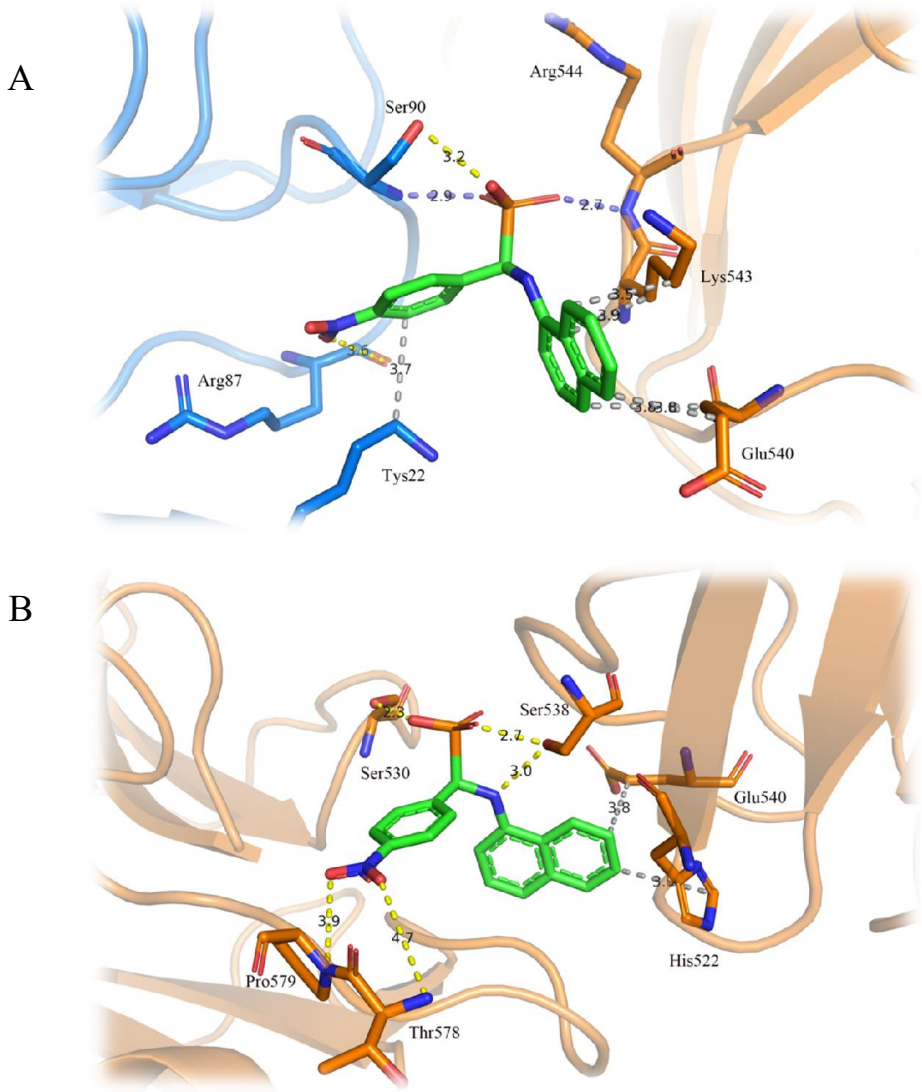


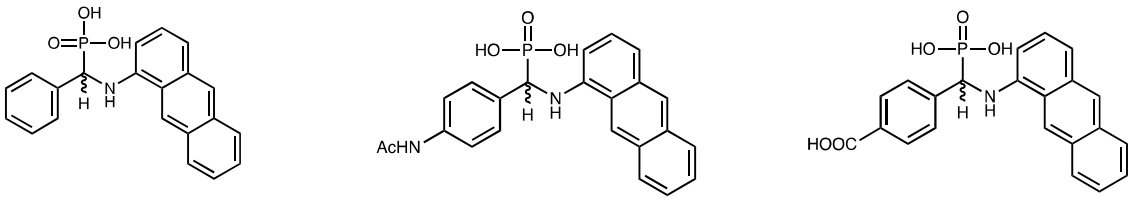
Figure 4.5. Details of interaction between compound 54 with TcTS in monoclinic (**A**) and orthorhombic (**B**) crystals.

Having the crystal structures in two different space groups indicates that the binding site between two SRM is not a real binding pocket but results from coincidental proximity of appropriate residues across a crystal contact. In these crystals, naphthalene-based inhibitors tend to be reside in the gap between two SRM, without major interactions with any key residues in the catalytic activity. In addition, no significant positive peak can be found to identify inhibitor molecules in crystals soaked with compound **53**. Thus, unfortunately, from these crystals, it is not possible to

determine the effective binding site of naphthalene inhibitors. All naphthalene compound are weak inhibitors, since they have low affinity or lower stability when interacting with TcTS, which appears to result in electron density that is not possible to identify in the real binding site.

4.5. The identification of a novel allosteric binding by crystal structure of TcTS complex with anthracene compounds

Among the novel aminophosphonate inhibitors, anthracene compounds present the best inhibition effect with a range of IC_{50} from 0.2 mM to 0.015 mM. All this type of inhibitor show non-competitive inhibition suggesting that they might bind to an allosteric site rather than the catalytic site, but the interaction of the inhibitor affects the catalytic mechanism without an impact on K_m value in present of inhibitor. The best candidate anthracene aminophosphonate inhibitors, compound **48, 59 or 61** (Figure 4.6)**,** was added into the drops in which the TcTS crystal would be soaked for 48 hours. X-ray diffraction data from all three crystals were collected at the synchrotron, and details of data collection and refinement are given in the table below (table 4.2).



48 **59** **61**

Figure 4.6. Structure of anthracene compounds soaked with TcTS crystals

Table 4.2. X-ray data collection and refinement statistics for crystals soaked with anthracene compounds

Crystal No.	5	6	7	8
Detector	In-house	Diamond beamline I03	Diamond beamline I03	Diamond beamline I04
Protein	TcTS			
Ligand soaked	48	48	59	61
Length of soaking	48 hour	46 hour	48 hour	45 hour
Ligand identified	48	48	59	61
Data collection and processing statistics				
wavelength (Å)	1.541	0.979	0.976	0.979
Resolution (Å)	65.06-1.69	54.26-1.31	48.19-1.39	51.72-1.43
Space Group	P 1 2_1 1	P 2_1 2_1 2_1	P 1 2_1 1	P 1 2_1 1
Unit cell dimensions				
a (Å)	54.36	90.51	54.38	54.28
b (Å)	130.11	119.37	129.72	129.48
c (Å)	54.64	60.92	54.63	54.47
α (°)	90	90	90	90
β (°)	108.42	90	108.19	107.64
γ (°)	90	90	90	90
Number of. unique observations	79536	147108	142388	126384
Redundancy	4.47 (3.13)	8.66 (3.02)	4.84 (3.38)	6.40 (4.25)
Completeness (%)	89.9 (81.1)	86.26 (78.5)	98.9 (90.7)	96.0 (67.2)
Rmerge (%)	0.050 (0.176)	0.046 (0.798)	0.049 (0.737)	03042 (0.799)
I/ σI	19.3 (5.1)	25.5 (1.5)	12.6 (1.2)	19.8 (1.4)
Refinement				
Reflections used	53147	116324	142349	126340
Number of protein residues	624	627	625	627
Number of water molecules	594	913	714	650
Number of atoms	10258	11916	11237	11016
R-factor	0.16	0.15	0.13	0.16
R-free	0.20	0.17	0.16	0.18
RMSD bond length (Å)	0.018	0.008	0.007	0.004
RMSD bond angle (°)	1.491	0.988	0.967	0.810

In the crystal structure soaked with anthracene compounds, a novel binding site was identified between the catalytic domain and lectin-like domain **(crystal 5),** and the distance of the anthracene compound binding site is about 30 Å from the catalytic pocket (Figure 4.7 A).

An orthorhombic crystal soaked with compound 48 **(crystal 6)** verified the position of this binding site (Figure 4.7 B), although they still have several minor differences. In monoclinic structures, the compound 48 binds to TcTS in only one conformation, but in orthorhombic crystal, a dual conformation can be identified from the electron density map (Figure 4.7 D), apart from the first conformation which is the same as the orthorhombic structure and other anthracene compounds (Figure 4.8A). However, a second conformation occurs in which the anthracene has a 90° rotation, and the phenyl group place different from the first one but still has cation-π iteration with Arg577 (Figure 4.8B).

Furthermore, in the orthorhombic crystal structure, a second binding site on the surface of the α-helix between the catalytic domain and lectin-like domain can be determined (Figure 4.7 E). It is possible that it is a crystal artifact since the same interaction cannot be found in either complexes of compound **48** with TcTS in different crystal forms or crystal soaked with different anthracene compounds. However, it might really has weak interaction because the soaked compound shows some interactions with residues on the helix, including the salt bridge between Ser380/Arg437 with phosphate group, and cation-π interaction between Arg437 and the anthracene group (Figure 4.8C). Furthermore, ITC analysis of anthracene compound binding indicated that the binding ratio between TcTS and this type of compound could be either 1:1 or 1:2 (ITC, n=1.3±0.25). Moreover, in both of two crystal forms the helix which links catalytic and the lectin-like domain is exposed to the solvent, result in the helical binding site is not affect by the symmetry related molecules. The helix binding site, thus, could be a real but weak binding site, it is because although the ligand shows interaction with residues on the helix, it only occur in the orthorhombic crystal form.

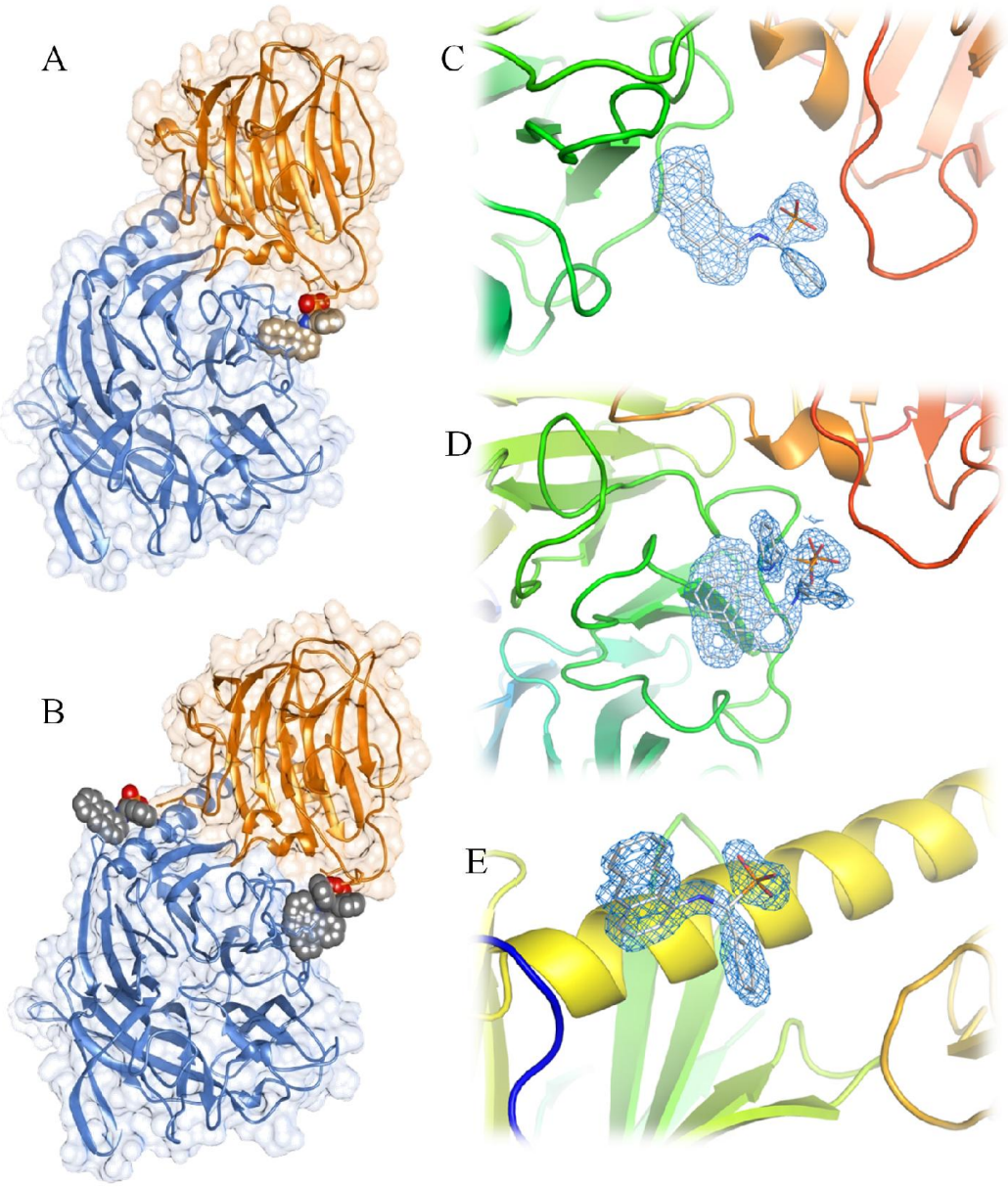


Figure 4.7. Binding location of compound **48** in monoclinic crystal (**A**) and orthorhombic crystal form (**B**). *2mFo-DFc* maps (σ=0.9) of compound **48** in monoclinic structure (**C**) and orthorhombic structure: **D** (Inter-domains binding site)**, E**(helix binding site).

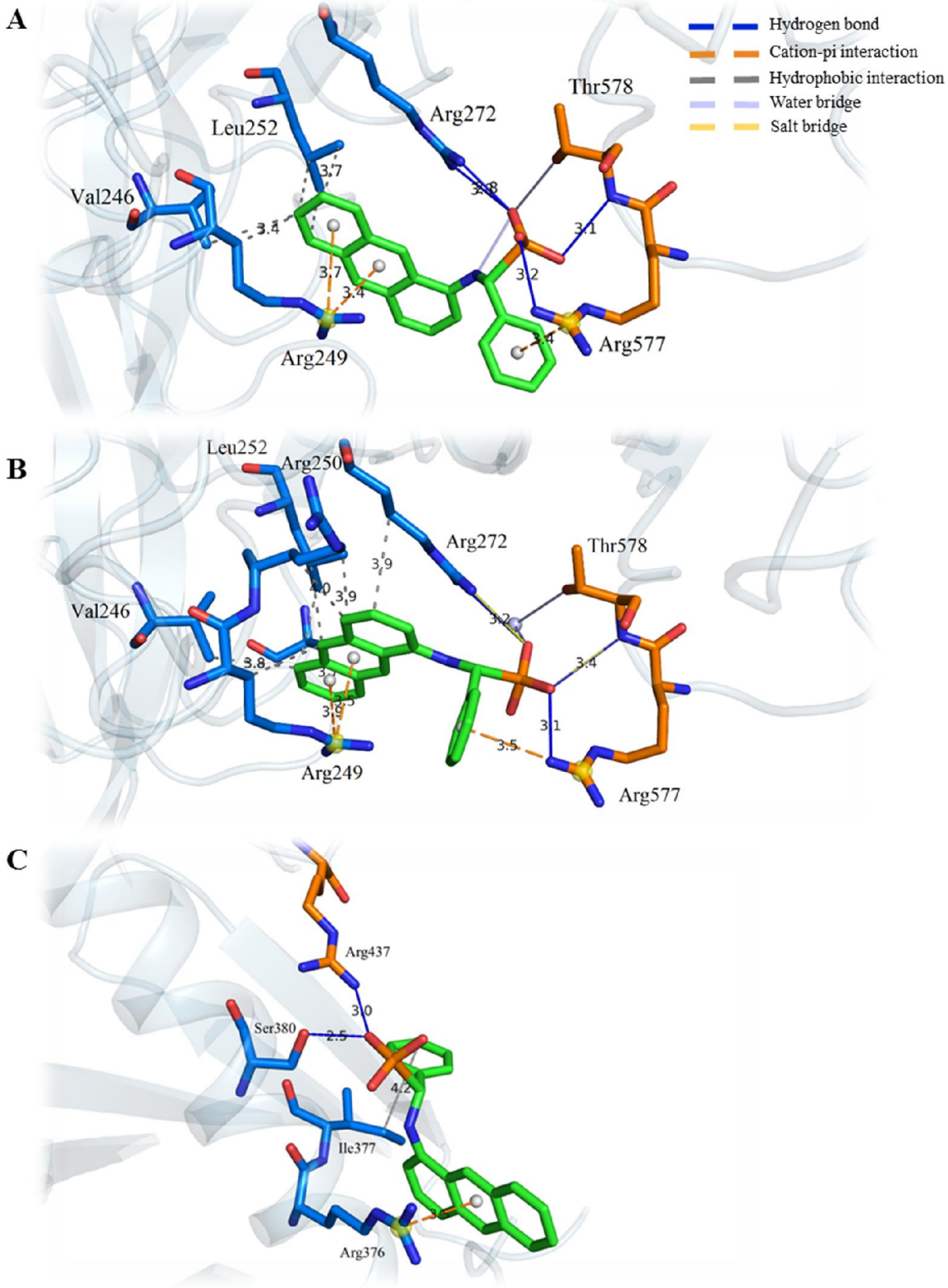


Figure 4.8. Interaction details of (**A**) First conformation, (**B**) second conformation of **compound 48** at inter-domain allosteric binding site and (C) helix binding site of **compound 48**. Ligands are present in green, catalytic domain residues are in blue and lectin-like domain residues are in orange.

Compound **59** (IC_{50} = 15µl) was the best inhibitor in the SAR screening of α-aminophosphonate derivatives, followed by the compound **61** (IC_{50} = 23µl). Crystal structures of these compounds soaked into TcTS **(crystal 7, 8)** showed that both bind to the inter-domain allosteric binding site identified by crystal structures of complexes with compound **48** (Figure 4.9). Thus, the anthracene compound soaks indicated that α-aminophosphonate derivatives, at least anthracene compounds, bind to TcTS in this novel binding site and it could be a potential target site for rational drug design of TcTS inhibitors.

Since anthracene compounds bind to the novel binding site in between two domains of TcTS, the interaction between ligand and TcTS is mediated by residues both from the catalytic domain (Arg249, Leu252, Arg272) and the lectin-like domain (Arg577, Thr578) (Figure 4.5). In all the structures with anthracene compounds bound, Arg249 and Leu252 clamp the anthracene moiety through cation-π interactions and hydrophobic interactions respectively. The phosphate group binds to Arg272 from catalytic domain, and Arg577 and Thr578 from the lectin-like domain. The amino nitrogen does not have direct interaction with TcTS but interacts with Thr578 via a water bridge. Compared with the non-liganded structure (Figure 4.9A), several residues in the inter-domains binding site change their conformation to interact with anthracene compounds. In the former TcTS structure, the side chain of Arg249 points towards to Tyr248, but on inhibitor binding a conformation change happens to form the cation- π interactionwith the inhibitor. Arg577 is s a flexible residue, which has often been truncated in non-liganded crystal structures (Buschiazzo. *et al.*, 2002).

Arg577 also undergoes significant movement to form a cation- π interaction with the phenyl group whilst also binding the phosphate group. In the interaction with aminophosphonate molecules, Arg249 and Arg577 act as a closing door to hold the inhibitor molecules into the binding site. As noted above, as well as with Arg577, the phosphate group of anthracene compounds also interacts with Arg272, the side-chain of which which shows a 52° rotation compared to the un-liganded TcTS structure. The

binding of aminophosphonate compounds result in fitting of binding site geometry might trigger the rotation of Arg272.

No structural data has been published in previous studies about the interaction of TcTS and either non-competitive or mixed inhibitors. Therefore, although numbers of non-competitive or mixed inhibitors have been developed based on structure mimicry of TcTS substrates, the allosteric binding site is still unknown.

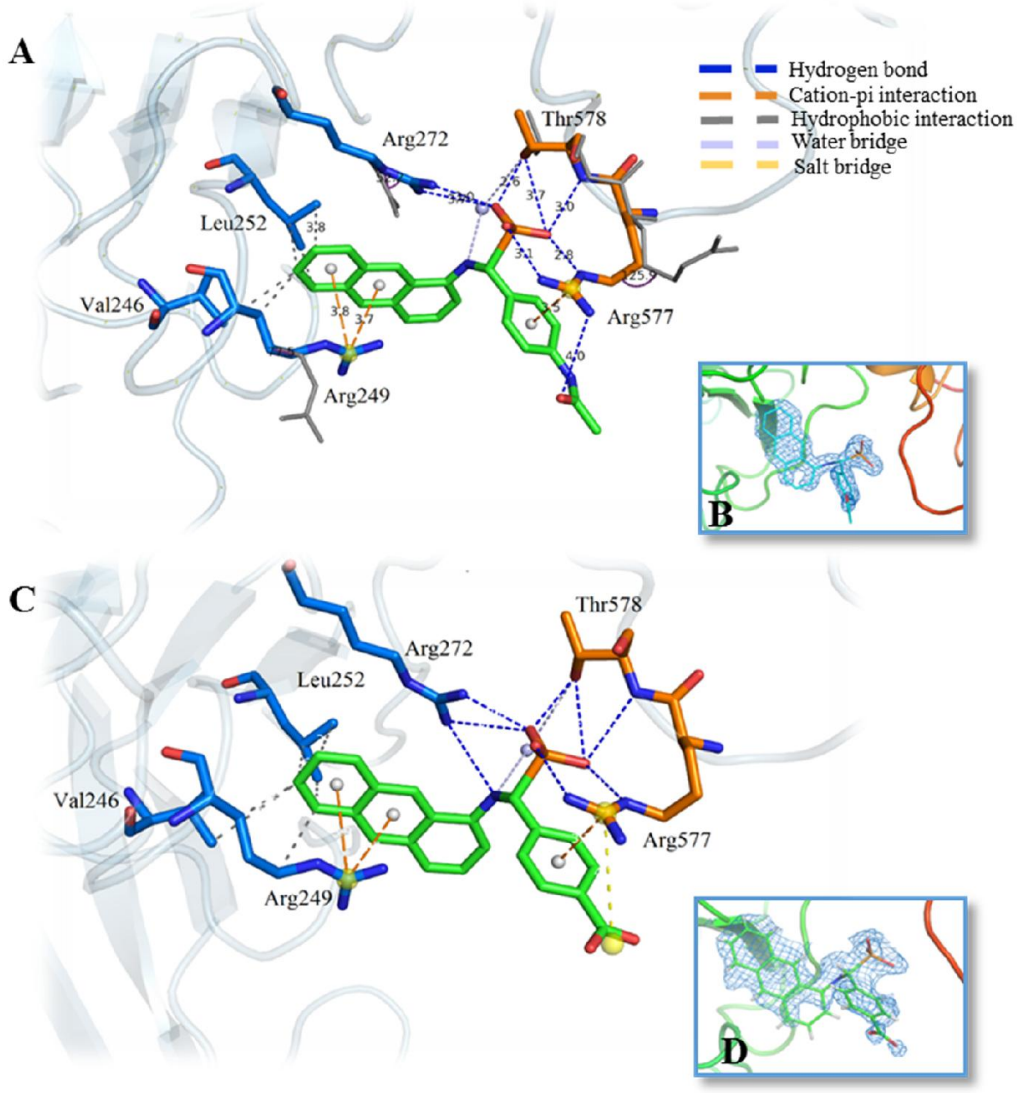


Figure 4.9 Interaction details, compound **59** (**B**) and **61** (**D**). Residues from catalytic domain and lectin-like domain are presented in green and orange respectively. Ligand molecules are in blue. *2mFo-DFc* maps (σ=0.9) of compound 59 and 61 are shown in **C** and **E**

The inter-domain binding site is located about 35 Å from the catalytic pocket. Therefore, whether it is a functional binding site and how this interaction will affect TcTS activity is important to understand for the next stage of study. The catalytic site in the crystal structures of anthracene inhibitors also shows several positive difference density peaks. However, they only models of components in mother liquor such as water, glycerol, PEG or PEG fragments fit into these, not the soaked inhibitor (Figure 4.10).

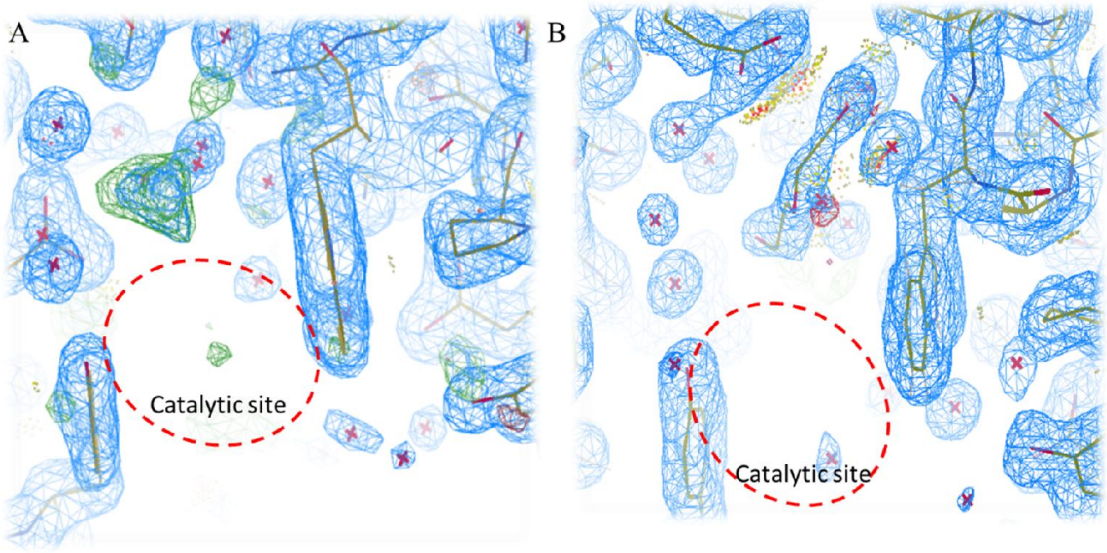


Figure 4.10. *2mFo-DFc* maps (blue, σ=1.5) and *Fo-Fc* map (Green for positive and red for negative, σ=3) of active site of TcTS soaked with **compound 59**, A: un-modeled, B: Model with PEG.

4.6. Validation of the allosteric binding site by introducing anthraquinone compounds

Crystal structures have indicated an allosteric binding site for TcTS interacting with α-aminophosphonate derivatives. However, when compounds with other aromatic groups such as naphthalene or phenyl at the "lactose-mimic" moiety were soaked into TcTS crystals, the binding position of ligand could not be identified due to the low protein-ligand affinity. Therefore, an anthraquinone group was introduced to validate the new binding site. Although anthraquinone derivatives have been shown to be decent inhibitors against TcTS, which has the range of IC_{50}s from lower millimolar level to micro molar level, and show non-competitive inhibition (Arioka *et al.*, 2010), the binding site and the inhibitory mechanism is still unclear.

Two compounds, compound **62** and **63**, containing anthraquinone group on the amino side of the inhibitor, were designed and synthesized (Figure 4.11). Inhibition screening results showed that anthraquinone compounds can inhibit TcTS activity more strongly than naphthalene or bi-aryl compounds but more weakly than anthracene compounds. Compound **62** with a phenyl group on the phosphate side has an IC_{50} of 94 μM and the IC_{50} of compound **63**, with a 4-accetamidophenyl group on the phosphate side is 79 μM (Figure 4.12 A, B; Table 4.3). Not surprisingly, they still show non-competitive inhibition against TcTS （Figure 4.12 C, D）

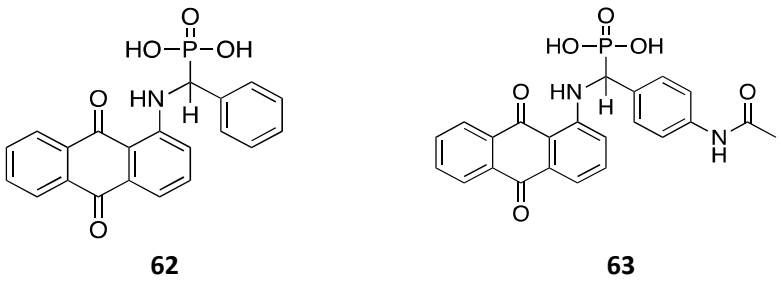


Figure 4.11. Structure of anthraquinone compound **62** and **63**

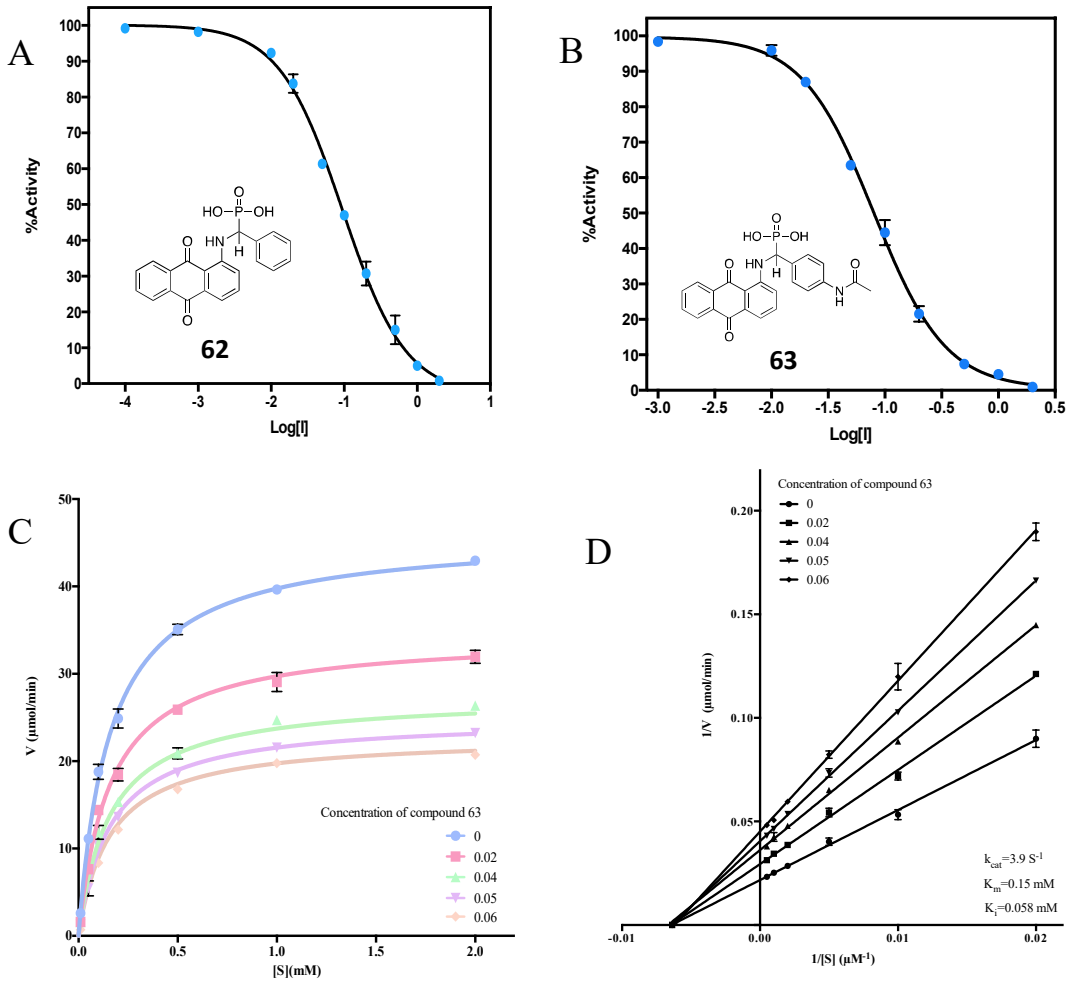


Figure 4.12. Inhibition and kinetics results for compound **62** and **63**

Table 4.3. Inhibition and kinetics result of compound **62** and **63**

Compound	$LogIC50_{TcTS}$	Std. error of $LogIC50_{TcTS}$	$IC50_{TcTS}$ (mM)	Inhibition mode	K_i (mM)
Compound 62	-1.024	0.027	0.094	-	-
Compound 63	-1.100	0.013	0.079	Non-competitive	0.058

In X-ray data collected from crystals soaked with compound **63** for 48 hours (**Crystal 9**), the compound can only be partially identified and built in the positive difference density at the inter-domain binding site (occupancy = 72%), and cannot be built in other positive difference map peaks. This crystal soaking result shows that compound **62** might still bind to the inter-domain binding site but the low occupancy and poor electron density map reflect the weaker affinity and lowered stability of the complex shown experimentally.

When compound **63** was soaked into TcTS crystals, the subsequent diffraction data revealed better density for the ligand in the inter-domain allosteric binding site since it is of higher affinity than **62**. Looking at the details of the binding (Figure 4.13 D), the interaction between anthraquinone compounds and TcTS is the same as for anthracene compounds. The phenyl group close to the phosphate still interacted with Arg577 through cation-π interaction, and the phosphate group interact with Arg577, Thr578 and Arg272. The anthraquinone group appears in between Arg249 and Leu252. However, in anthracene compound soaked structures, the cation center of the Arg249 is around 3.6-3.8 Å above the middle and the outer ring of the anthracene, however, in anthraquinone compounds, Arg249 only has cation-π interaction with the outer ring since the conformation changes.

In crystal structures of either naphthalene or anthracene compounds, there is no inhibitor bound to the catalytic site. However, electron density which fits the structure of an anthraquinone fragment has been defined in the structure soaked with compound **63** (Figure 4.13 C). The anthraquinone fragment undergoes π-π stacking with both Tyr119 and Trp312, and is located in a similar place as the lactose fragment in the SAL soaked structure (Figure 4.13 E).

Table 4.4. X-ray data collection and refinement statistics for TcTS crystals soaked with anthraquinone compounds

Crystal No.	9	10
Detector	In-house	Diamond beamline I04
Protein	**TcTS**	
Ligand soaked	**62**	**63**
Ligand identified	**62**	**63**
Data collection and processing statistics		
wavelength (Å)	1.541	0.979
Resolution (Å)	65.06-1.69	52.24-1.58
Space Group	P 1 2_1 1	P 1 2_1 1
Unit cell dimensions		
a (Å)	54.36	54.39
b (Å)	130.11	129.82
c (Å)	54.64	54.88
α (°)	90	90
β (°)	108.42	107.83
γ (°)	90	90
Number of unique observations	79536	91114
Redundancy	8.84 (4.47)	6.84 (4.47)
Completeness (%)	89.9 (81.1)	92.2
Rmerge (%)	0.050 (0.176)	0.048(0.035-1.011)
I/ σI	19.3 (5.1)	19.2(1.2)
Refinement statistics		
Reflections used	53147	91076
Number of protein residues	624	623
Number of water molecules	594	320
Number of atoms	11356	10043
R-factor	0.16	0.18
R-free	0.20	0.23
RMSD bond length (Å)	0.018	0.014
RMSD bond angle (°)	1.491	1.235

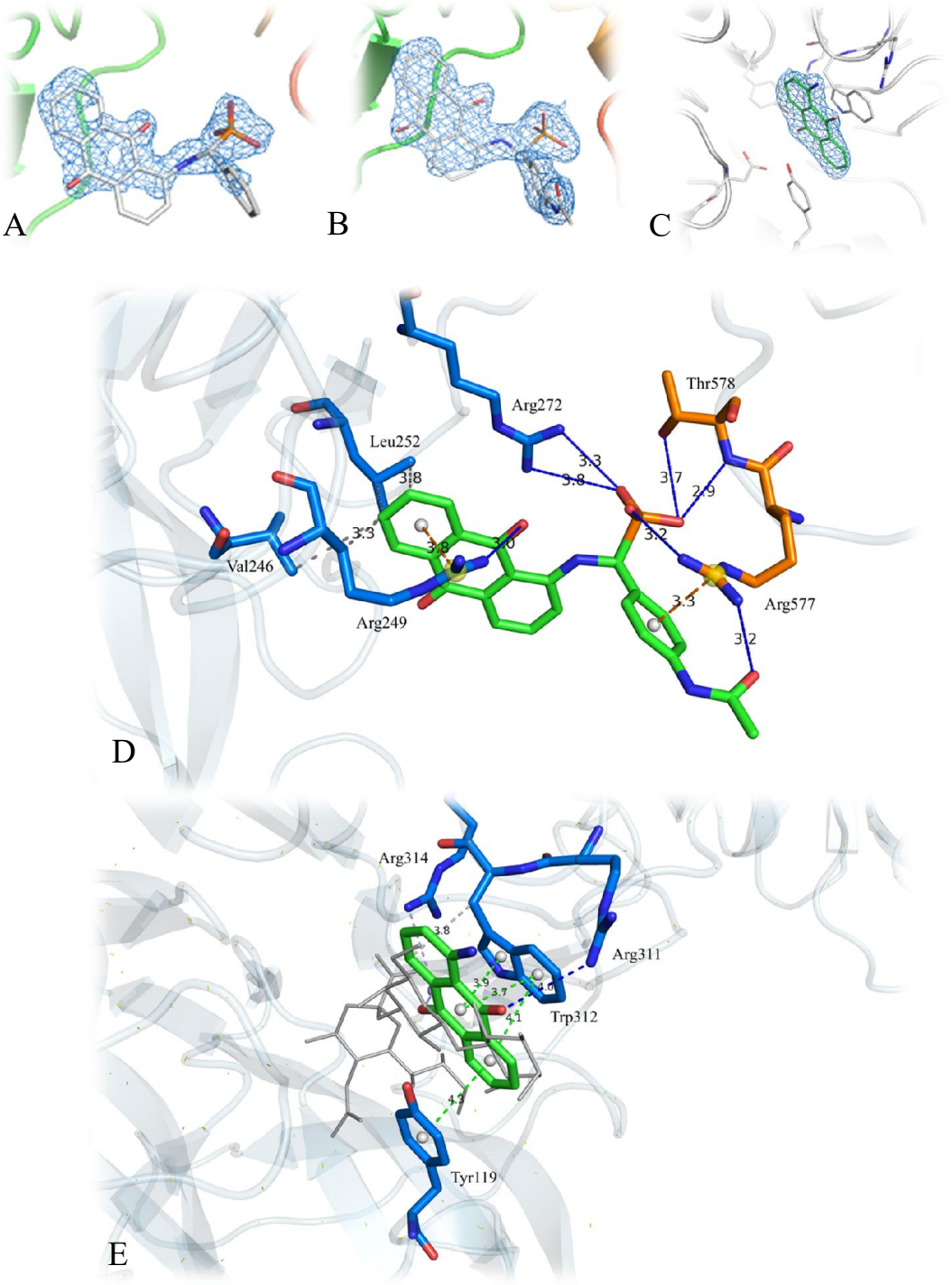


Figure 4.13. *2Fo-mFc* map of compound **62** (**A**) and **63** (**B**) in the allosteric site. *2Fo-mFc* map of anthraquinone fragment of compound **63** in lactose binding site (σ = 0.9). **C**: Details of interaction between compound 62 at the allosteric site. **D**. Details of iterations of anthraquinone in the lactose binding site, superimpose with SAL (PDB: 1S0I).

4.7. Crystal structure of published inhibitors (DANA, lactitol, rhein)

As well as the α-aminophosphonates, other published inhibitors were also soaked into TcTS crystals and their X-ray crystal structures determined, including DANA, lactitol, and rhein. DANA was the first defined inhibitor against TcTS, although it only shows extremely weak inhibition (Vandekerckhove *et al.* 1992). It was reported that after soaking in DANA for short time, TcTS crystals were destroyed as the result of active site conformation changes (Buschiazzo *et al.* 2002). In this study, we show that in the monoclinic crystal form, the crystal survived after 48 hours soaking and provided high resolution (1.7Å) and high completeness diffraction data (table 4.5).

Table 4.5. X-ray data collection and refinement parameters of DANA, lactitol and rhein soaked into TcTS crystals

Crystal No.	11	12	13
X-ray source	In-house	In-house	Diamond beamline I03
Protein	TcTS	TcTS	TcTS
Ligand soaked	DANA	Lactitol	Rhein
Length of soaking(hours)	48	48	48
Ligand identified	DANA	N/A	Rhein
Data collection and processing statistics			
X-ray wavelength (Å)	1.541	1.741	0.979
Resolution (Å)	1.69 (54.23-1.69)	1.78 (54.23-1.78)	1.59 (64.82-1.59)
Space Group	P1 2_1 1	P1 2_1 1	P1 2_1 1
Unit cell dimensions			
a (Å)	54.38	54.32	54.52
b (Å)	130.18	130.02	129.65
c (Å)	54.62	54.65	54.82
α (°)	90	90	90
β (°)	108.54	107.33	107.43
γ (°)	90	90	90
Unique Number of. reflections	75397	76644	96484
Redundancy	4.22 (2.93)	4.84 (2.66)	6.32 (2.86)
Completeness (%)	93.8 (79.3)	88.7 (65.3)	99.22
Rmerge (%)	0.095 (0.232)	0.092 (0.365)	0.0964
I/ σI	9.2 (2.5)	6.2 (1.2)	7.2 (2.3)

Refinement			
Reflections used	74684	71738	96444
Number of protein residues	624	623	623
Number of water molecules	633	656	634
Number of atoms	10276	10150	10055
R_{factor}	0.16	0.16	0.21
R_{free}	0.18	0.19	0.24
RMSD bond length (Å)	0.005	0.019	0.010
RMSD bond angle (°)	0.787	1.491	1.037

When compared to the published crystal structure of TcTS with DANA, when DANA is soaked into the crystals for a long time, the structure shows similar interaction between the enzyme and ligand, indicated that binding of DANA is stable in the catalytic site and not affected by the time of interaction. Unlike in the structures from crystals soaked with α-aminophosphonate compounds, in the DANA binding structure Tyr119 is present with 100% occupancy in the "open" conformation, and interact with oxygen O9 at the glycerol side chain (Figure 4.14). Although DANA is only a weak inhibitor of TcTS, the (*2Fo-mFc*) electron density surrounding the DANA is much better defined and appears in higher occupancy then the stronger inhibitors such as α-aminophosphonate and rhein ($occ._{DANA}$=0.93 and $occ._{Rhein}$=0.76). This phenomenon could be explained by DANA buried inside the catalytic pocket that has higher conformational stability than the allosteric binding site of the α-aminophosphonates, or the lactose-binding site where Rhein binds, which are on the surface of the enzyme and might be affected by higher exposure level to the solvent.

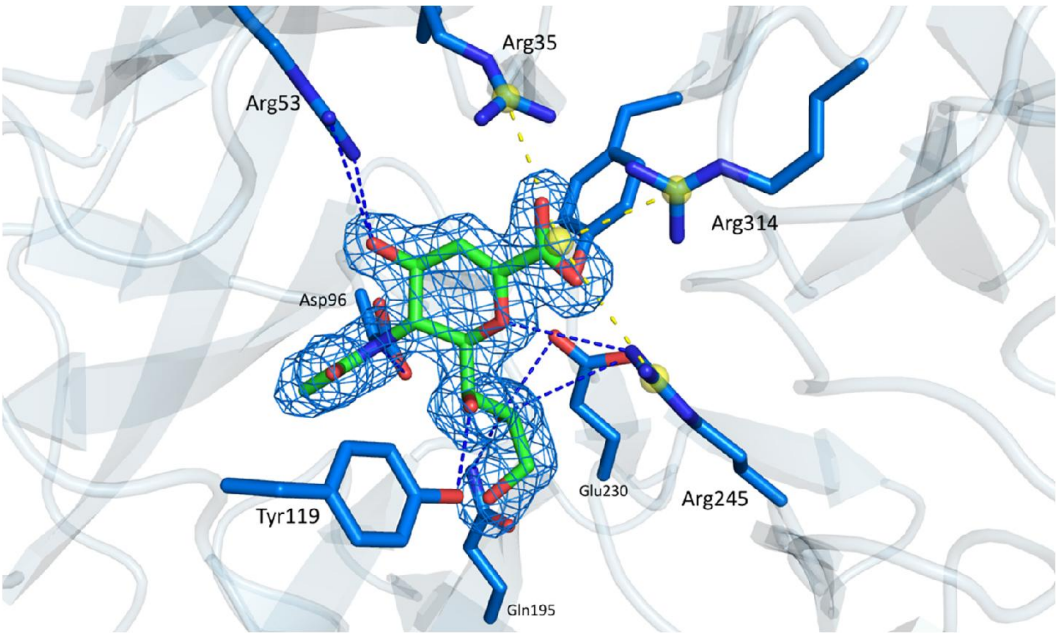


Figure 4.14. Crystal structure of DANA interacts with the TcTS catalytic site, with *2mFo-DFc* map of DANA molecule (σ=0.9, occ. = 0.93). Hydrogen bonds are present by blue dash and salt bridge by in yellow dash.

Lactitol is the only molecule on the market that can inhibit TcTS with an IC_{50} lower than 1mM (Agusti, Paris *et al.* 2004), and it has been applied in research into TcTS mechanism and functions (Mucci, Risso *et al.* 2006). However, there is still no crystal structure available for lactitol bound to TcTS, though it is thought to interact with TcTS in a similar manner to lactose binding. A TcTS crystal was soaked with a solution with up to 10mM lactitol for 48 hr, and 1.7 Å diffraction data was obtained from the in-house X-ray system. After molecular replacement and several rounds of refinement, lactitol was able to be identifed at the lactose-binding site (Figure 4.15), indicating that lactitol binds to TcTS similarly to lactose. However, the weak occupancy (occ.=73.2%) suggests lactitol is a weak binder to TcTS, but it still can bind to the lactose binding site without presence of a sialic acid donor. In the lactitol-soaked structure, Tyr119 has moved to the parallel conformation, which indicates the formation of the lactose binding site, the cyclic ring of lactitol appearing in the stacking position between Tyr119 and Trp312 but it did not have strong interactions, shown by the long distances from lactitol to both Tyr119 (4.6 Å) and Trp312 (4.7 Å).

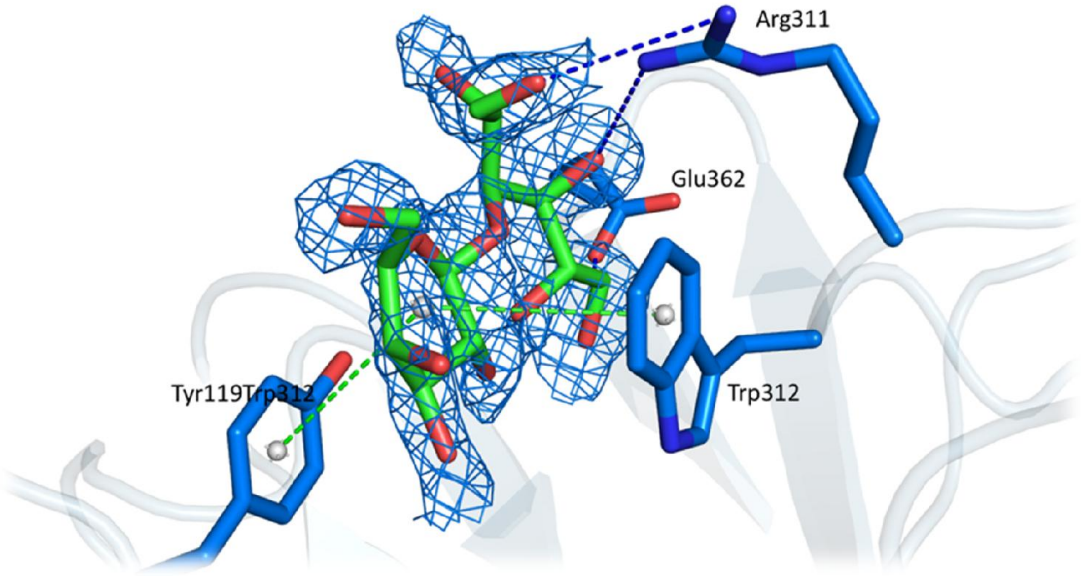


Figure 4.15. Detail of interaction of lactitol at the TcTS lactose binding site, with *2mFo-DFc* map of lactitol molecule (σ=0.7, occ. = 0.73). Hydrogen bonds are present by blue dash and π- π stacking is shown by green dash.

Rhein is an anthraquinone compound which was identified in 2010 by Arioka *et, al.* that can inhibit TcTS activity more strongly than DANA and lactitol (IC_{50}=17μM) and it also shows non-competitive inhibition. (Arioka *et al.* 2010). However, the binding site of anthraquinone compound is still undefined.

The crystal structure of TcTS in complex with anthraquinone-aminophosphonate above (Section 4.3.5) showed that as well as the binding in the inter-domain allosteric binding site, positive difference electron density was also defined in the lactose binding site between Tyr119 and Trp312 (Figure 4.16) into which an anthraquinone fragment could be built. To investigate whether anthraquinone compounds can bind to the lactose-binding sites, a TcTS crystal was soaked in solution (condition B) containing 0.18mg/ml TcTS and 0.2mM of rhein. As expected, following data collection and structure solution, positive difference map peaks can be identified in the lactose binding site and fit the model of rhein molecule (Figure 4.16).

Compared to the anthraquinone-aminophosphonate compound, the interaction between rhein and TcTS lactose binding site shows stronger π- π stacking, and all three of the aromatic rings are stacked with either Tyr119 or Trp312, and this could be the reason that rhein exhibits stronger inhibition of TcTS than anthraquinone-aminophosphonate compounds. In addition, since the lactose binding site for SA transfer activity is unique to trypanosoma trans-sialidase rather than other sialidases, it may explain the selectivity of anthraquinone scaffold.

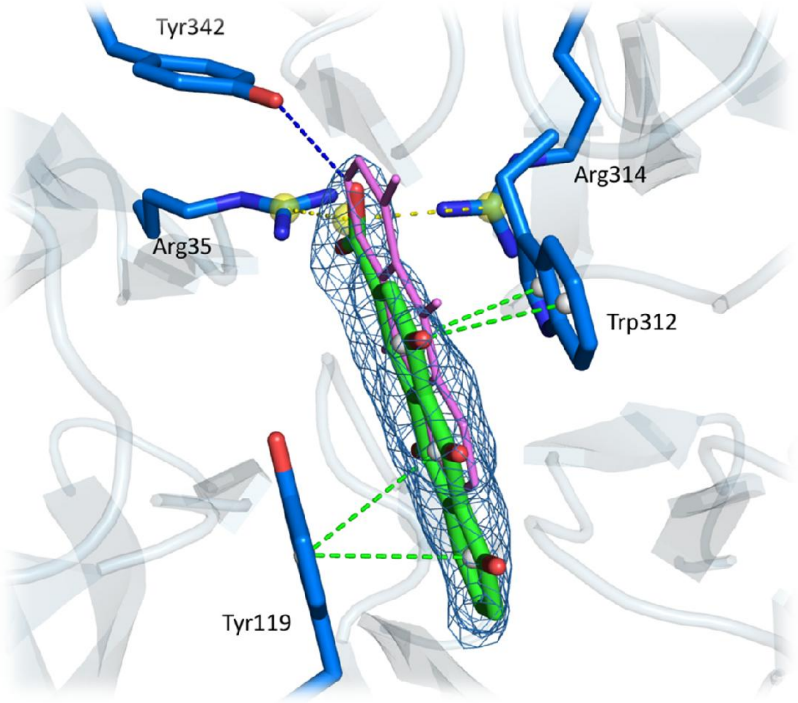


Figure 4.16. Crystal structure of rhein interaction at the TcTS lactose-binding site, with *2mFo-DFc* map on rhein molecule (σ=0.9, occ.=0.86). Superimposed with anthraquinone fragment of compound **63** at the catalytic site (pink). Hydrogen bond is present by blue dash, salt bridge by in yellow dash, and π- π stacking is by green dash.

4.8. Binding of α-aminophosphonate affects the conformation of the catalytic site

As non-competitive inhibitors, it can be observed from the crystal structures that α-aminophosphonate derivatives bind to TcTS at an allosteric pocket between the catalytic and lectin-like domains. Therefore, it is important to explain the inhibitory mechanism of this allosteric binding. The allosteric site location is about 35 Å from the catalytic site and several residues in the catalytic domain, such as Arg249, Leu252 and Arg27, participate in the interaction with aminophosphonate compounds. However, these residues have not been recorded to be involved in the TcTS catalytic activity. Therefore, the effect of the allosteric binding on the catalytic activity or even the catalytic site structure cannot be directly estimated from existing reports. At the catalytic site, it can be identified that Tyr119, appears in the conformation that stacks with Trp312, which known as the "closed position".

Tyr119 has been shown to be a critical residue for the transfer activity of TcTS, loss of Tyr119 will significantly impact both hydrolysis and transfer activity (Paris *et al.*, 2001), and gain of tyrosine in the same position in *T. rangeli* sialidase leads to this enzyme acquiring the ability to transfer sialic acid (Paris *et al.*, 2005). Generally, Tyr119 is rotated towards the catalytic pocket in empty structures or in an intermediate state interacting with sialic acid, which are regarded as the "open" conformation. The Tyr119 stacking conformation, also known as the "closed" conformation, normally occurs in secondary ligand binding, for example, binding with donor or acceptor oligosaccharides (Amaya *et al.*, 2004; Buschiazzo. *et al.*, 2002; Watts *et al.*, 2006). However, the closed conformation also occurs in empty TcTS structures without ligand binding, indicating the high flexibility of Tyr119 (Buschiazzo. *et al.*, 2002; Telford, 2014).

To investigate whether the binding of allosteric inhibitor triggers the closed conformation of Tyr119, co-soaking of both compound **48** (5 mM) and DANA (10 mM)

into a TcTS crystal was performed. X-ray data collection and refinement information is given in Table 4.6

Table 4.6. X-ray data collection and refinement of TcTS crystal soaked with compound **48** and DANA

Crystal No.	14
Detector	Diamond beamline I02
Protein	TcTS
Ligand soaked	DANA+**48**
Length of soaking(hr)	48hr
Ligand identified	DANA, **48**
Data collection parameters	
X-ray wavelength (Å)	0.979
Resolution (Å)	51.56-1.23
Space Group	P1 2_1 1
Unit cell dimensions	
a (Å)	54.27
b (Å)	129.58
c (Å)	54.31
α (°)	90
β (°)	108.3
γ (°)	90
Number of unique reflections	189537
Redundancy	
Completeness (%)	91.36 (86.45)
Rmerge (%)	0.044 (0.031-1.190)
I/ σI	
Refinement parameters	
Reflections used	180253
Number of protein residues	626
Number of water molecules	724
Number of atoms	10788
R-factor	0.1532
R-free	0.1684
RMSD bond length (Å)	0.017
RMSD bond angle (°)	1.466

The previous DANA-soaked crystal structure (Section 4.3.6) showed that DANA is a good catalytic site binder although it is only an extremely weak inhibitor against TcTS. In the co-soaked structure with both DANA and compound **48**, compound **48** was identified at the inter-domain binding site, and DANA in the catalytic pocket (Figure 4.17 A) where it is bound in the same position as in the TcTS-DANA complex structure (Figure 4.17 B). This indicates that binding of α-aminophosphonate compounds will not impact the assessment and interaction of substrate or other catalytic site binder. However, in the co-soaked TcTS structure, Tyr119 presents a dual conformation that has a ratio between open and closed occupancies of 3:7. This result shows that the allosteric binding of α-aminophosphonate inhibitor tends to change the conformation of Tyr119 from the open to closed state (Figure 4.18), that is an important step of lactose binding site formation. The lack of a positive difference map peak at the lactose-binding site stacking between Tyr119 and Trp312, means that the "closure" of Tyr119 might not because of lactose site occupancy. Although the Tyr119 still has a higher occupancy at open than the closed position, this may be due to DANA and the α-aminophosphonate compounds not being strong inhibitors. Therefore, it can be postulated that a possible inhibitory mechanism of the α-aminophosphonate compounds in the allosteric site is a result of the conformational change of Tyr119 and formation of stable lactose binding site, result in lactose binding without donor substrate and prevent it from release.

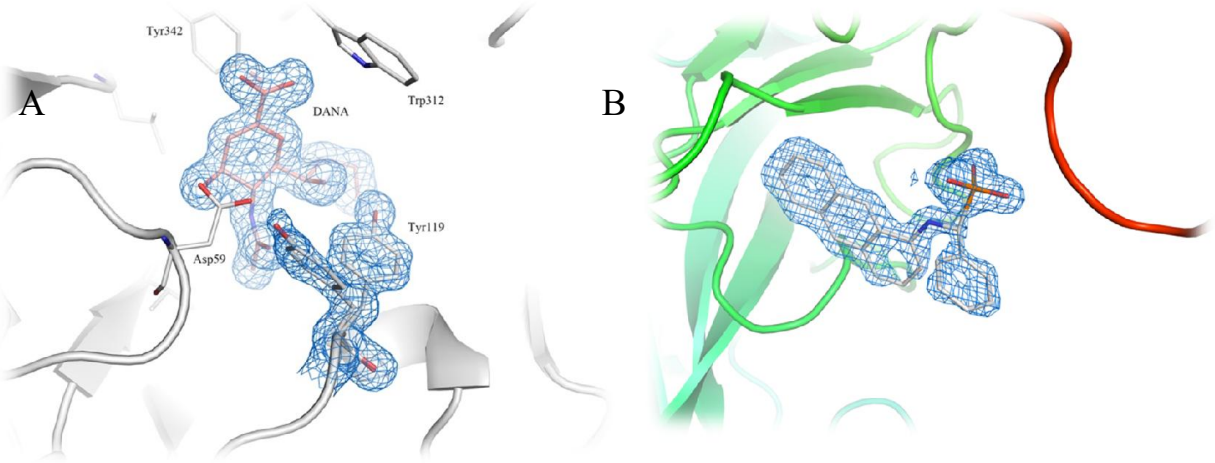


Figure 4.17 (**A**): The allosteric site showing the model of compound 48 in the DANA/48 co-soaked TcTS structure with the *2Fo-DFc* map (σ=1.0) drawn around the 48. (**B**) The catalytic

site of TcTS in the co-soaked structure with f DANA and the dual conformations of Tyr119 surrounded by the 2*Fo-DFc* map (σ=1.0).

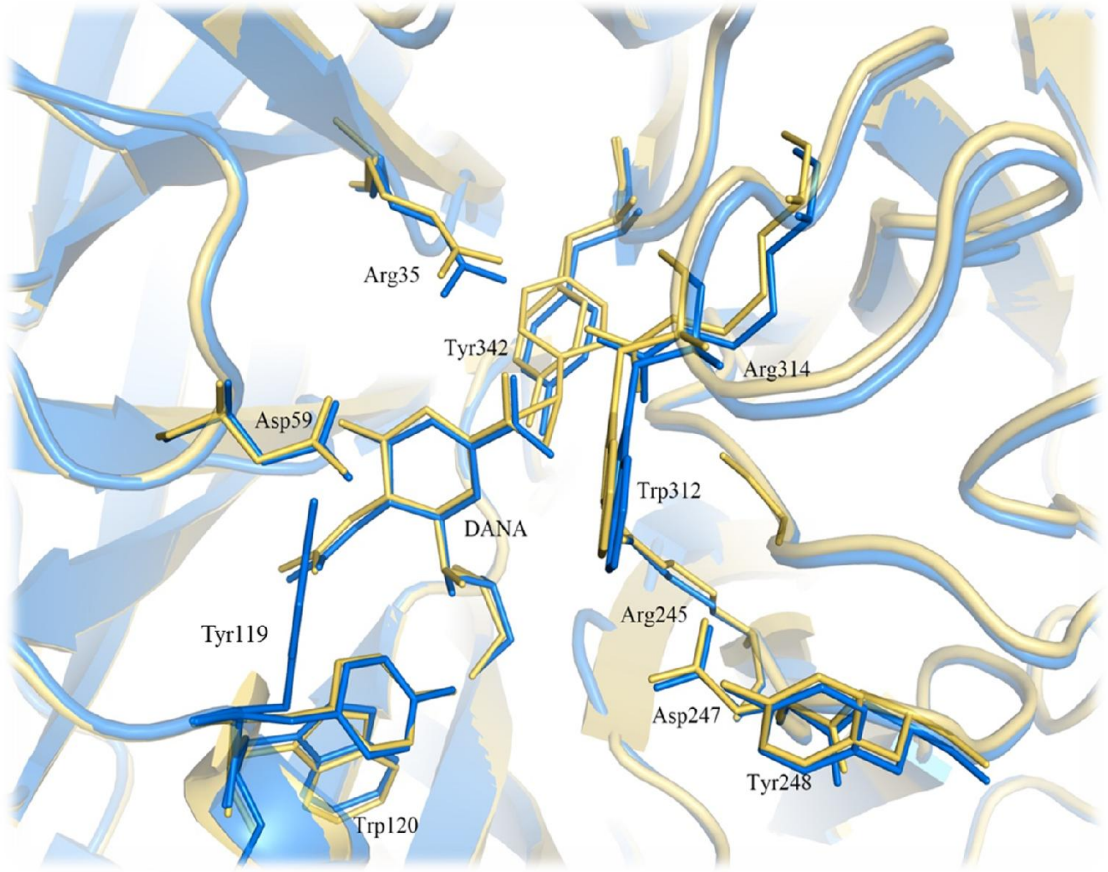


Figure 4.18 Structure alignment of the catalytic pockets of the co-soaked crystal with DANA and compound 48 (Blue, **crystal 14**) and that of the DANA soaked TcTS crystal (Yellow, **crystal 11**).

4.9. Mutagenesis study on the TcTS allosteric binding site

As known from crystal structures of TcTS in complex with α-aminophosphonate, it can be hypothesed that the interaction at the inter-domains allosteric binding site could change the conformation of the catalytic site, leading to inhibition. Since the residues in the allosteric binding site are distant from the catalytic site, they will not be directly involved in the catalytic activity, but they might also control the catalytic activity through indirect means such as influencing substrate recognition.

To investigate the effect of allosteric binding site on catalytic activity, site-directed mutagenesis on the three arginine residues at the allosteric site was performed. As shown in the crystal structures (Figure 4.9), the most important residues interacting with aminophosphonate compounds are three arginines, Arg249, Arg272 in the catalytic domain and Arg577 in the lectin-like domain. As seen in Figure 4.19, the overall activity was not eliminated by any of the mutations, showing that none of these three residues is critical for TcTS activity. The turnover number, which is represented by $k_{cat.}$ has slightly reduced in all three mutations from 4.2 s^{-1} to 3.9 s^{-1}. Surprisingly, even the interaction with α-aminophosphonate inhibitor did not change the K_m of substrate but mutation of any of the inhibitor-binding residues may decrease the substrate affinity, with an increasing K_m from 0.15mM to 0.24-0.25mM. The kinetic results indicate that the allosteric binding site could have a function in TcTS activity and may be involved in substrate recognition.

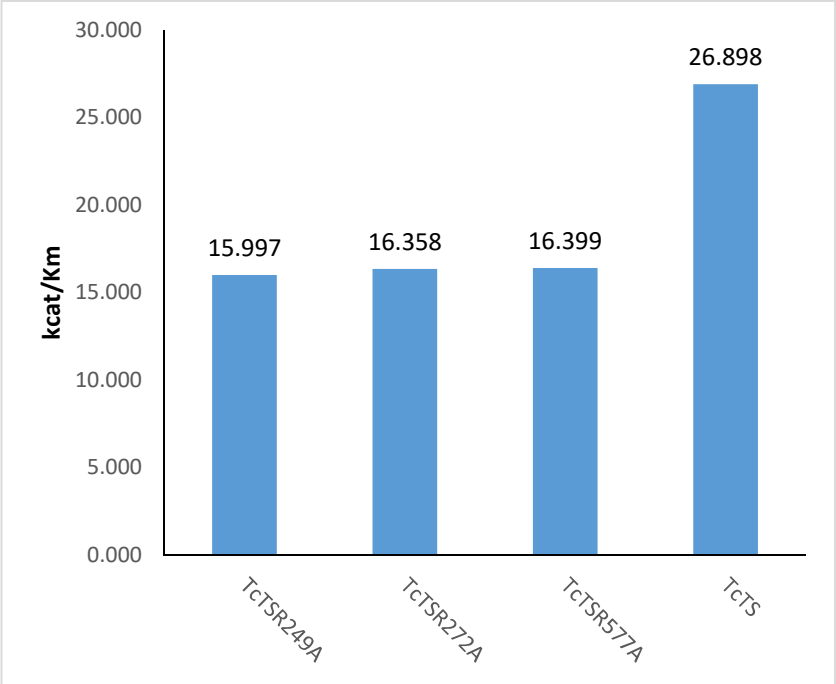


Figure 4.19 k_{cat}/K_m of WtTcTS and TcTS mutants in allosteric site (R249A, R272A, and R577A)

Table 4.7 Kinetic parameters of TcTS mutations compared to wild type TcTS

	$TcTS_{R249A}$	$TcTS_{R272A}$	$TcTS_{R577A}$	*TcTS*
k_{cat} (s^{-1})	3.95	3.92	3.94	4.18
K_m (*mM*)	0.25	0.24	0.24	0.16
k_{cat}/K_m	15.99	16.36	16.40	26.87

4.10. Sequence alignment study of allosteric binding site

The variation of TcTS between different strain is the main reason why TcTS cannot be a potential vaccine target (Rosenberg *et al.*, 2016). To investigate whether the α-aminophosphonate compounds could be a broad-spectrum inhibitor or a strain specific inhibitor against TcTS, multiple sequence alignments (MSA) were undertaken to compare TcTS protein sequences within different isoforms. The proteome of the *T. cruzi* CL brener strain has been intensively studied (Sodre *et al.*, 2009), thus, sequences of TcTS isoform can be obtained from the CL brener strain for MSA study. The similarity between TcTS isoforms from CL brener strain and the query sequence (Q6966) is shown in table 4.8, sequence XP_807431.1 has the highest identity and the sequence XP_80548.1 has the lowest identity, which is 77%.

The MSA result (Figure 4.12, full alignment in Appendix 8.7) shows that most inhibitor binding residues are conserved across TcTS isoforms, suggesting that the allosteric binding site interacting with α-aminophosphonate compounds is generally conserved among different TcTS strains. However, Arg249 only occur in two different isoforms and is replaced by alanine and threonine in others. Interestingly, in the isoforms containing Arg249, residue 248 is tyrosine (XP_807308.1 and XP_807593.1), however if residue 249 is alanine, polar residues, such as arginine appear at residue 248, which might play a similar role to Arg249 in interactions with inhibitors.

Table 4.8. The sequence accession codes of TcTS from CL brener strain and statistics of similarity produced by BLAST (Altschul, *et al*., 1990) when searched with the query sequence (Q26966).

TcTS Accession Code	Max score	Total score	Query cover (%)	E value	Maximum Identity (%)
XP_807431.1	1281	1281	99	0.0	96
XP_806228.1	1270	1270	99	0.0	95
XP_816523.1	1259	1259	99	0.0	95
XP_807499.1	1204	1204	99	0.0	91
XP_807074.1	1194	1194	98	0.0	91
XP_807593.1	1134	1134	99	0.0	87
XP_813615.1	1129	1129	98	0.0	86
XP_807308.1	1096	1096	99	0.0	85
XP_815267.1	1092	1092	98	0.0	85
XP_806454.1	1004	1004	96	0.0	78
XP_804548.1	978	978	96	0.0	77

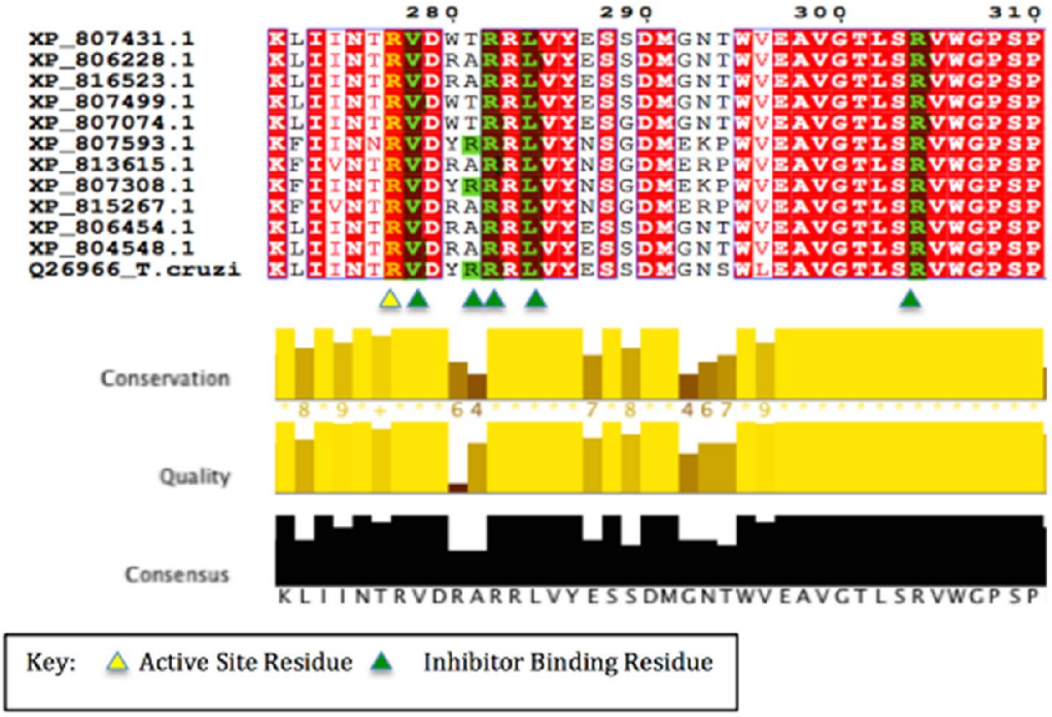


Figure 4.20. The inter-domain binding site region of MSA result of different TcTS isoform. Key residues at the inter-domain binding site are highlighted in green and catalytic site residues are in yellow

4.11. Discussion

4.11.1. TcTS crystal structure shows the non-competitive inhibition of α-aminophosphonate compounds

Acquiring a protein-ligand complex is the critical step for structure-based drug design and development. For TcTS, although numbers of potential types of inhibitor have been identified, the lack of a structure-function relationship study will limit further development of these inhibitor, especially for inhibitors which do not bind to the known binding sites. The case of the α-aminophosphonate and other examples above shows that although compounds may be designed based on the substrate of an enzyme, they may not bind to the enzyme competitively. Fortunately, the α-aminophosphonates can still interact with TcTS and inhibit its activity in a non-competitive manner. The crystal structures of TcTS-aminophosphonate complexes can attest the non-competitive inhibition through the novel inter-domain binding site and lack of electron density for inhibitors in the catalytic site. Through kinetic assays (chapter 3) it has been shown that the binding of an α-aminophosphonate compound does not affect the binding K_m of the donor substrate; this has also been proved by the co-soaked TcTS structure showing both the aminophosphonate inhibitor and DANA bound. In general, allosteric binding is accompanied by significant conformation change of the acceptor (Luque and Freire, 2000). However, from the soaking experiments to obtain crystals of TcTS in complex with aminophosphonate inhibitors, it is clear that, no major conformational change occurs in the overall structure of TcTS, and looking at the structure, even the catalytic site is almost the same as the unliganded structure. Therefore, it is difficult to explain the inhibition mechanism based on inhibitor soaking structures.

An anthraquinone group was introduced at the AM moiety to validate the allosteric binding site. The anthraquinone compounds have weaker inhibition than the anthracene

compound, and have lower occupancy in the crystal structure. This can be explained by the introduction of two C=O bonds changing the electrostatic distribution which weakens the cation-pi interaction with Tyr249 (Figure 4.21).

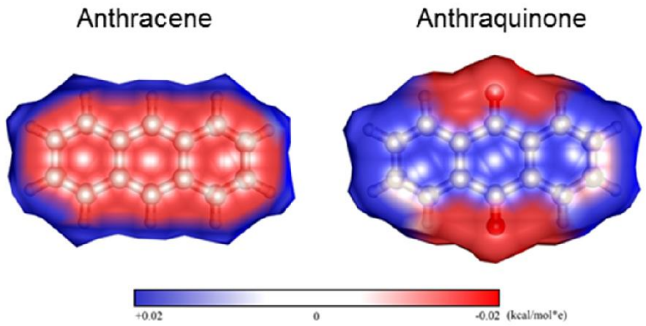


Figure 4.21. Electrostatic distribution of anthracene and anthraquinone (generated by UCSF Chimera) with positive charges in blue, negative in red

Through allosteric binding, the inhibition mechanism of aminophosphonate inhibitors could be attributed to three possible reasons: (1). Binding of an aminophosphonate inhibitor could change the conformation at the active site. (2). although the interaction of this type of inhibitor cannot block the substrate binding, it could inhibit the release of catalytic product by increasing the rigidity and stability of the active site. (3). the inter-domain binding site has a function in the catalytic activity such as substrate recognition.

From crystal structure of TcTS-aminophosphonate complexes, no significant conformation change can be observed in the catalytic pocket except the Tyr119, which flips to the “closed” position. However, since Tyr119 is present in a dual conformation in non-liganded structure, it is still not possible to decide whether the closed conformation is result of inhibitor binding. Thus, the addition of DANA to the crystal soaking was performed since DANA is known to be a good active site binder and can trigger the rotation of Tyr119 to the open position. Comparison of the DANA-soaked and aminophosphonate-soaked TcTS structures showed that although the binding of aminophosphonate inhibitor cannot fully induce the conformation change of Tyr119, it

can still partially trigger the formation of the closed conformation, which is critical for lactose binding.

In addition, the anthracene or anthraquinone moiety of aminophosphonate compounds interact with a loop which contain several important residues in TcTS activity, for example Asp247 (neighboring residues 248 and 249 involved in aminophosphonate binding) interacts with Trp312 via a water bridge, which could trigger the conformation change of Trp312 away from Tyr119 (Figure 4.22). Mutation of Asp247 will significantly reduce TcTS activity (Oliveira *et al.*, 2014). In unliganded TcTS structures, this loop is exposed to the solvent and tends to have high flexibility to affect the movements of Trp312. By interacting with aminophosphonate inhibitors, this loop might become more stable and the conformational changes of Trp312 could be restricted, which will affect the binding and release of substrates.

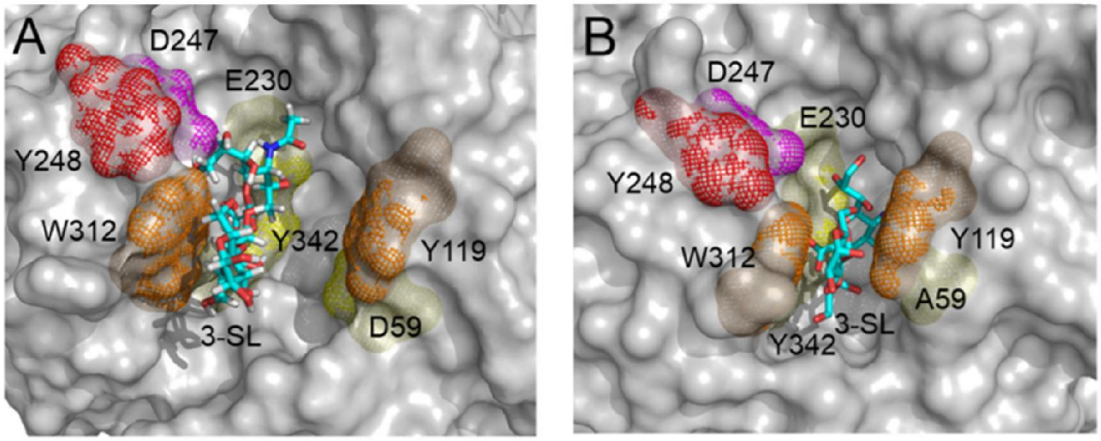


Figure 4.22. A: final frame of the MD simulation of binary complex TS3-SL. **B**: crystal structure of intermediate complex of $TcTS_{D59A}$-SAL (Oliveira *et al.*, 2014).

The R249A mutation did not eliminate or significantly reduce TcTS catalytic activity, but the K_m and k_{cat} for MuNANA substrate were slightly different in the mutant enzyme. However, this result is insufficient to show that the inter-domain binding site is critical for TcTS activity.

4.11.2. Inhibitors probing lactose binding site show non-competitive inhibition

The anthraquinone derivatives are good inhibitors for the trans-sialidase, but after incorporation into aminophosphonates, the performance of anthraquinone seems weaker than the anthracene they replaced. The reason could firstly be attributed to the weaker cation-π interaction between anthraquinone compounds and TcTS. Rather than being parallel with the aromatic group as in the example of anthracene compound binding, the Arg249 is attracted to the oxygen in the anthraquinone group and loses the optimal distance and angle to from the strong cation-π interaction (Marshall *et al.*, 2009). Another reason could be that multiple binding sites weaken the interaction at the inter-domain binding site, since an anthraquinone fragment can be found in the lactose binding site and the rest of compound seems placed outside the catalytic site, suggesting that around the catalytic site, the only stable interactions can be made by the anthraquinone fragment.

The less modified anthraquinone derivatives also show very weak inhibition. For example, Dantron (Figure 4.23A) has the IC_{50} against TcTS at 0.71 mM, but the addition of a carboxyl group at C2 (rhein, Figure 4.23B) improves the IC_{50} value to 0.014 mM (Arioka *et al.*, 2010), and the carboxyl group also has great contribution to interactions based on the crystal structure (Figure 4.16). Thus, it can be expected that the anthraquinone fragments at the aminophosphonate compounds (Figure 4.23C) might act weakly as a anthraquinone inhibitor that interacts with the lactose binding site. The weaker inhibition at both the inter-domain binding site and lactose binding site results in weaker inhibition of anthraquinone aminophosphonate compounds compared to anthracene compounds.

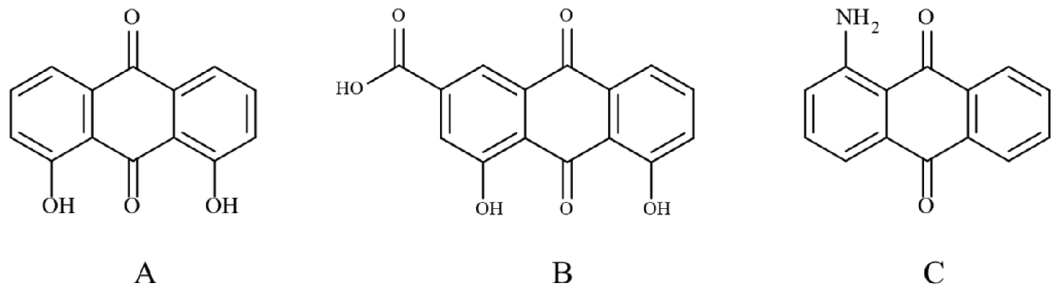


Figure 4.23. Structure of **(A)** Dantron, **(B)** Rhein and **(C)** anthraquinone fragment of **compound 62** which identified in the lactose binding site.

Either in published anthraquinone derivatives (Arioka *et al.*, 2010) and anthraquinone-aminophosphonate compounds are non-competitive with MuNANA, and interact with the lactose-binding site, which is part of the catalytic pocket. These results suggest that that when the lactose-binding site is occupied by anthraquinone, the donor substrate can still access to the SA binding site. Comparing with the position of anthraquinone and DANA/SA binding (Figure 4.24), it can be seen that after interacting with anthraquinone at the lactose-binding site, it might still have enough space for donor sialiyl substrate to bind to the SA pocket and be hydrolyzed, but the anthraquinone will block the binding of acceptor substrate. That could be an explanation for the non-competitive inhibition for the lactose site binder.

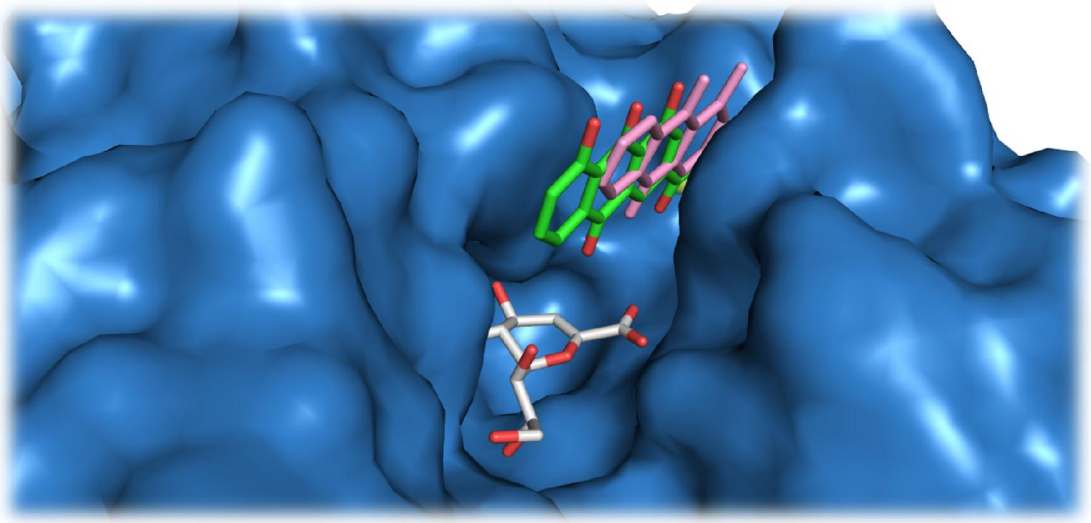


Figure 4.24. Binding of anthraquinone in the lactose binding site, superimposed with crystal structure with DANA binds with the SA binding site. Rhein is shown as green stick, anthraquinone fragment of compound 62 is pink stick, and DANA is molecule is present as white stick.

Both structure and kinetics studies suggest that the lactose binding site of TcTS is formed by donor substrate interaction, that can explain why lactose derivatives or other acceptor substrates can only interact with the lactose binding site whilst the catalytic pocket is occupied by sialic acid (Buschiazzo. *et al.*, 2002). Agusti *et al* suggested that lactitol inhibits TcTS activity as a poor acceptor substrate but also a good lactose-site binder (Agusti *et al.*, 2004). When TcTS crystals are soaked in a high concentration of lactitol, it can be identified in the catalytic site, although with a low occupancy. But at the lactose binding site of crystal structure soaked with lactose, no lactose molecules could not be identified (Buschiazzo. *et al.*, 2002), suggesting that lactitol has higher affinity than lactose and is able to bind to TcTS without donor substrate being present.

The crystal structure of TcTS in complex with rhein suggests that rhein or other anthraquinone compounds interact with the lactose binding site without binding of sialic acid, similarly crystal structures of TcTS in complex with anthraquinone-aminophosphonate compounds also indicated that the anthraquinone moiety can bind to the lactose binding site. However, due to the effect of the phosphonate and phenyl groups on the aminophosphonate compound, only two rings of the anthraquinone stack with Tyr119 and Trp312, and all three aromatic rings show π- π stacking with lactose binding site residues, which induces a stronger interaction with the lactose binding site than anthraquinone-aminophosphonate compounds.

As a known non-competitive inhibitor, binding of anthraquinone derivatives, would not affect the binding TcTS substrates, even though the lactose-binding site has been occupied. It suggests that it might have extra binding site for the oligosaccharide moiety of the donor sialyl oligosaccharides. That could be an experimental evidence to support the hypothesis that the TcTS mechanism is not a traditional ping-pong but a mixture of a ping-pong and ternary structure model (Oliveira *et al.*, 2014).

4.11.3. Conclusion

Using structural biology approaches, the high-resolution structures of TcTS in complex with different ligands have been determined. The structural studies revealed a novel binding site between the catalytic domain and lectin-like domain. Crystal structures also suggest lactitol and anthraquinone bind to the lactose-binding site, providing new understanding of TcTS inhibition mechanisms that the lactose site binding of inhibitors can be also non-competitive to the donor substrate. It can also correlate the relationship between ligand affinity and identification of ligand in crystal structures, weak inhibitors might be difficult to identify.

Chapter 5. Interpreting lactose binding: evaluation of the inhibition mechanism of TcTS by α-aminophosphonates

5.1. Prelude

Despite their design based on the structure of the TcTS substrate, α-aminophosphonate compounds have been determined to be non-competitive inhibitors and it was shown to bind to an allosteric binding site between the catalytic and the lectin-like domain due to the determination of the crystal structure of TcTS-inhibitor complex. However, the mechanisms of inhibition of the allosteric binding are still unclear. The information deduced from the crystal structures also needs confirmation using other biophysical or biochemical approaches.

The detailed mechanisms of TcTS, especially the binding of the acceptor lactose and the transfer activity, are still controversial (Damager *et al.*, 2008; Oliveira *et al.*, 2014). In our studies, formation of the lactose binding site is achieved in by interacting with α-aminophosphonate in the TcTS-inhibitor complex structures, therefore, lactose binding can be used to investigate the mechanisms of inhibition by α-aminophosphonate derivatives. Additionally, lactose binding behavior is also useful in understanding the mechanisms of TcTS catalytic activity.

This chapter describes the investigation of lactose binding to TcTS in the presence of inhibitor, using enzyme kinetics, structural biology, and thermodynamic assays. These studies interpret the influence of lactose on TcTS catalytic activity, and its effects in present of α-aminophosphonate. The crystal structure of TcTS co-soaked with both inhibitor and lactose suggests that the formation of lactose binding site through α-aminophosphonate binding truly occurs, and this result is also confirmed by the thermodynamic studies. Altogether, the mechanism of TcTS inhibition by α-aminophosphonates can be attributed to formation of the lactose binding site, and induction of a stable interaction with lactose or other acceptor oligosaccharides. In addition, binding of the α-aminophosphonate also limits the release of catalytic product from the catalytic site, thus inhibiting the activity of TcTS.

5.2. Effect of lactose on kinetics and inhibition

In the TcTS catalytic activity, lactose or other oligosaccharides play an important role as acceptor substrates or as part of donor substrates. Therefore, the influence of lactose is critical in any investigation of TcTS inhibition. In this study, MuNANA was used as the substrate to detect TcTS hydrolysis activity since the catalytic product 4-Mu is neither an inhibitor of TcTS nor the acceptor substrate, thus only hydrolysis activity would be detected. In the presence of lactose, the kinetics characteristics of the hydrolysis activity might be affected due to the interposition of acceptor substrate and transfer activity.

In the same reaction system, both the K_m and k_{cat} of MuNANA were significantly increased when high concentrations of lactose (5mM) were added into the reaction mixture, indicating that in the presence of acceptor substrates, the K_m of MuNANA is reduced whilst the turnover number has increased (Table 5.1). This can be explained since with the addition of acceptor substrates, TcTS shows trans-sialidase activity, an extra step of sialyation onto lactose occurs that influences the interaction with SA, inducing the reduction of K_m of donor substrate. Moreover, because of the attack on the TcTS-SA intermediate by lactose, the SA moiety becomes easier to release from the catalytic pocket through sialylation activity, the turnover time is reduced and the turnover number of TcTS then increases from about 3.8 ± 0.3 s^{-1} to 7.6 ± 0.2 s^{-1}. Additionally, the k_{cat}/K_m ratio did not show significant changes, suggesting that the catalytic efficiency remains at the same level.

Table 5.1 Enzyme kinetics of TcTs in the presence of inhibitor **compounds 53 and 48** with or without lactose

Compound	*Lactose (5mM)*	K_m(mM)	k_{cat}(s^{-1})	k_{cat}/K_m	IC_{50}(mM)	K_i (mM)
48	-	0.15	3.6	24.0	0.048	0.035
	+	0.34	7.8	22.9	0.042	0.032
53	-	0.15	4.2	28.0	0.23	0.20
	+	0.31	7.5	24.2	0.21	0.24

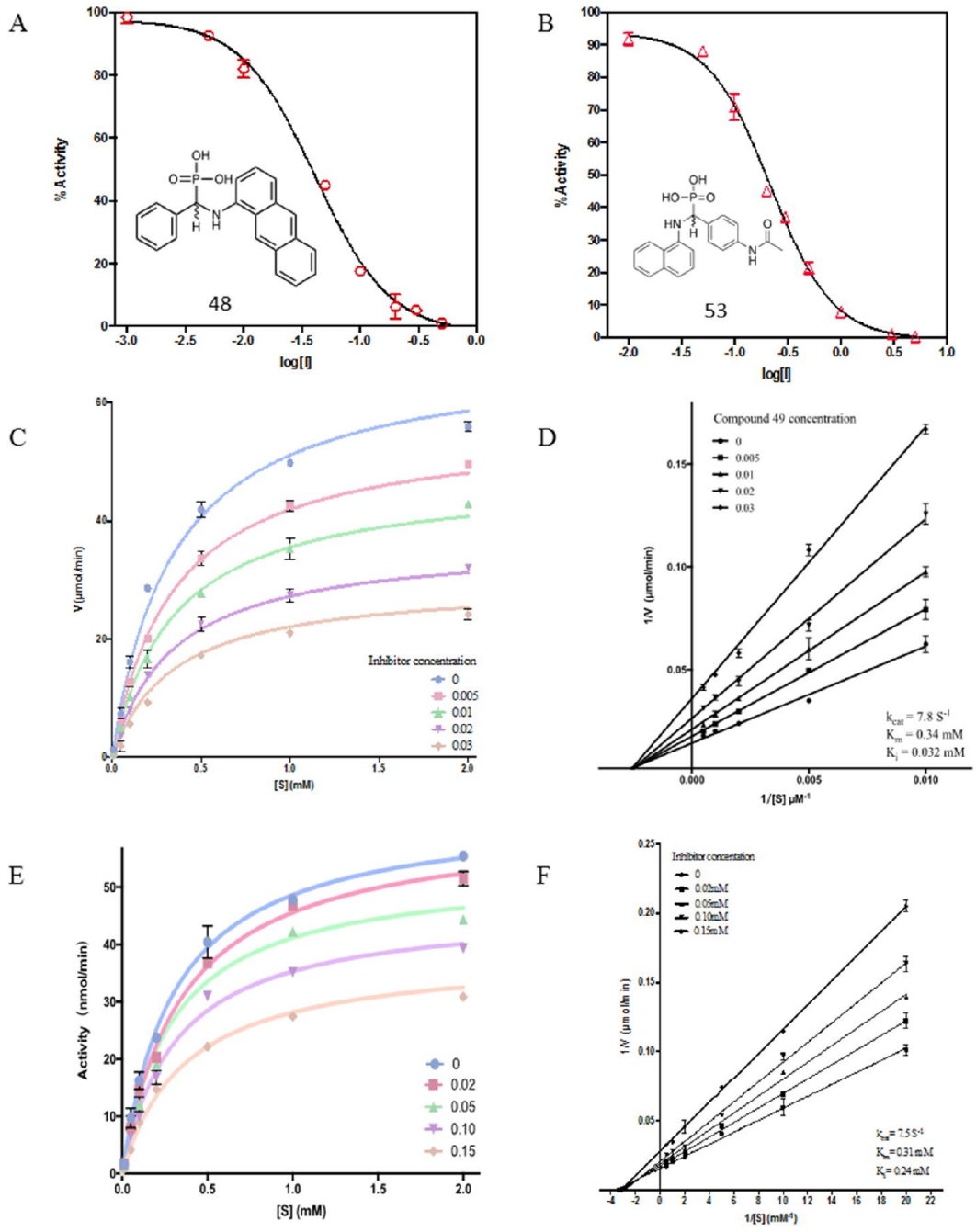


Figure 5.1 A: Non-linear fitting of IC_{50} determination of compound **48**; **B**: Non-linear fitting

of IC_{50} determination of compound **53**; **C, D.** Non-linear fitting and Lineweaver-Burk plot of TcTS kinetics in the presence of compound **48** in different concentrations. **E, F.** Non-linear fitting and Linewaver-Burk plot of TcTS kinetics in the presence of compound **53** in different concentrations.

Inhibition assays without lactose show that the inhibition of α-aminophosphonate follows non-competitive behavior, suggesting that the interaction between TcTS and this type of compound would not affect the binding of substrates, MuNANA in this case. The allosteric binding site which is observed in crystal structures indicates that α-aminophosphonates do not interact with either the SA or lactose binding sites. Therefore, the influence of lactose binding in the presence of inhibitors has been investigated in this study.

Both IC_{50} and inhibition assays were performed with 5 mM of lactose, which is much higher than the K_m value of lactose in the presence of SA (0.79 mM, using MuNANA as donor substrate) (Wilbrink, M. H. *et al.*, 2014), to prevent depletion of lactose during the assay. The IC_{50} of both detected inhibitors show slightly reductions in the presence of lactose but these are not significant, indicating that binding of lactose has no or limited influence on the binding and inhibition of TcTS by α-aminophosphonate compounds. The addition of an acceptor substrate did not change the mode of inhibition, and the inhibition constant K_i of both compounds investigated was unchanged.

These results support that, from an enzymatic point of view, α-aminophosphonates do not bind to the catalytic site, nor to the lactose moiety directly. Additionally, the addition of inhibitor does not influence the binding of lactose, both the K_m and k_{cat} value still increase when inhibitor interacts with TcTS. However, the kinetic assay was based on the fluorescence released from the hydrolysis product 4-Mu. Thus, the donor substrate, which can trigger the lactose binding site formation, is crucial for the kinetic assay. However, it is still unclear whether the inhibitor can induce lactose binding site formation. And how the lactose is being processed during the catalytic process,

especially in the sialyation step, cannot be determined in this assay. Therefore, further investigations were carried out to determine whether lactose binding with TcTS in the presence of the α-aminophosphonate inhibitors can be informative in the interpretion of the inhibition mechanisms.

5.3. Crystal structure of lactose bound to TcTS

A single crystal of TcTS was grown as described in Chapter 4, then transferred to crystallization solution (condition B, chapter 4.3.2) containing 18 mg/ml purified TcTS and 5 mM compound 48 and incubated at room temperature for 16-20 hours. The soaked crystal was then transferred to crystallization solution (condition B) containing 18 mg/ml purified TcTS, 5 mM compound 48 and 10 mM lactose. Crystal soaking with lactose was also performed in condition B with 10 mM of lactose, i.e. without inhibitor. The soaking crystals were incubated at room temperature for 12 hours, then cryo-cooled using by liquid nitrogen. X-ray diffraction data collection was performed using synchrotron radiation (Diamond Light source, UK).

In the presence of lactose, the hydrolysis efficiency of TcTS is maintained at the same level with the increase of k_{cat} and decrease of K_m. The presence of lactose cannot change either the effect or mode of inhibition of α-aminophosphonates. In crystal structures of TcTS soaked with α-aminophosphonate compounds, the formation of lactose binding site was observed. Structure and surface plasmon resonance (SPR) studies show that lactose only binds to TcTS in the presence of bound sialic acid or its derivatives, suggesting that the lactose binding site only forms when the sialic acid binding pocket is occupied (Buschiazzo. *et al.*, 2002). Therefore, whether the interaction of α-aminophosphonate compounds can induce the lactose binding site formation and allow lactose binding without the presence of sialic acid or other donor substrates is critical in interpreting the mechanisms of inhibition and the mechanisms of the TcTS catalytic activity. Binding of α-aminophosphonate compounds can induce the rotation of Tyr119

to the “closed” position, which indicates the formation of the second binding site that allows either donor or acceptor oligosaccharides to bind. The TcTS crystal structure with DANA and compound **48** (4.3.7) also indicates that Tyr119 can still rotate toward the sialic acid binding pocket when it is occupied by sialic acid or its derivatives, but the binding of aminophosphonate compounds tend to retain the conformation of Tyr119 in the closed state. Therefore, whether binding of aminophosphonate inhibitors can trigger lactose binding in the lactose binding site is crucial to interpreting the inhibitory mechanism of aminophosphonate derivatives.

Soaking of TcTS crystals in lactose solution without SA or other donor substrates had no electron density for lactose in the lactose-binding site (Buschiazzo. *et al*., 2002), this same result was obtained in our lab after the crystal had been soaked in 10 mM lactose for as long as 72 hours (data not shown). Nevertheless, using high resolution data (1.4 Å, table 5.2) from a TcTS crystal co-soaked with both compound **48** (5mM) and lactose (10mM), lactose can be identified in the catalytic site.

Table 5.2 Data collection and refinement of TcTS crystal soaked with compound **48** and lactose (crystal structure 15)

Crystal No.	15
Beamline	Diamond beamline i03
Protein	TcTS
Ligand soaked	10 mM Lactose 5 mM **compound 48**
Length of soaking (hours)	48
Ligand identified	Lactose and compound 48
X-ray data collection	
X-ray wavelength (Å)	0.976
Resolution (Å)	51.28-1.37
Space Group	P1 2_1 1
Unit cell dimensions	
a (Å)	51.28
b (Å)	128.76
c (Å)	54.06
α (°)	90
β (°)	108.45
γ (°)	90
No. unique reflections	189537
Redundancy	6.7 (5.9)
Completeness (%)	93.0 (60.7)
Rmerge (%)	0.046 (1.017)
Refinement	
Reflections used	135949
Number of protein residues	625
Number of water molecules	689
Number of atoms	10934
R-factor	0.16
R-free	0.17
RMSD bond length (Å)	0.005
RMSD bond angle (°)	0.811

Since TcTS has low affinity for lactose even in the presence of sialic acid (K_m = 0.79 mM), a higher concentration of lactose was used in the crystal soaking to enhance lactose binding (Vandekerckhove *et al*., 1992). As reported previously (Buschiazzo. *et al*., 2002), a TcTS crystal structure soaked with lactose alone failed to define any lactose molecules or fragments. However, in the co-soaked crystal structure of TcTS with both lactose and compound **48**, after molecular replacement and a few cycles of refinement, both of the soaked ligands can be identified interacting with TcTS (Figure 5.2). Electron density map peaks which fit compound 48 can be found in the inter-domain allosteric binding site as described before, and the lactose molecule can also be identified in the lactose binding site. After refinement, the electron density map for the lactose is still poorer than that of the inhibitor ligand, with an occupancy of 78%, which is slightly lower than either the protein (occ. = 1) or the inhibitor compound **48** (occ. = 0.83), suggesting that either lactose has a weaker binding with the lactose binding site, which tends to be formed by the interaction between TcTS and the inhibitor, or that the conformation of lactose has a lower certainty.

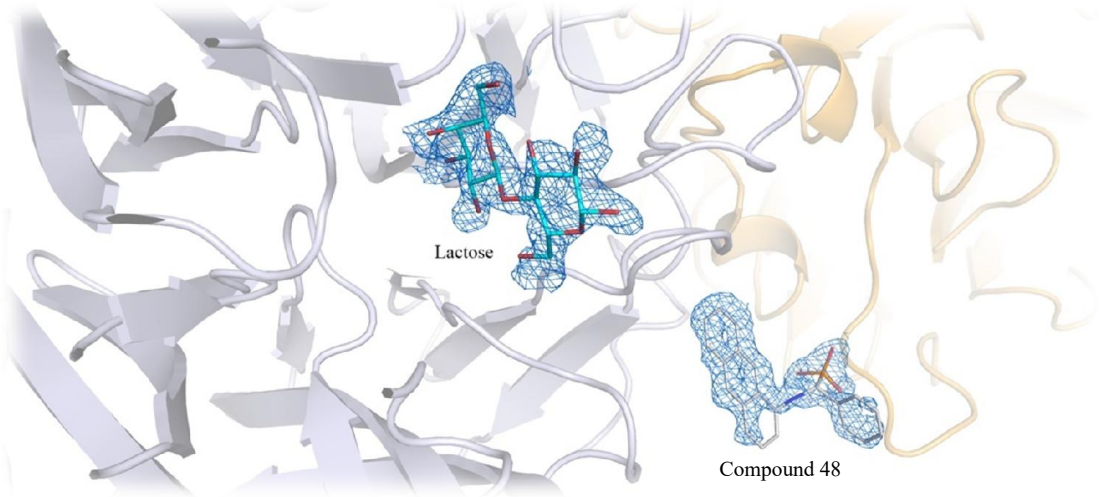


Figure 5.2. The electron density map (*2Fo-mFc* map contoured at σ = 0.8) of lactose and compound 48 (in cyan and grey stick form respectively) bound in the TcTS crystal structure (shown in cartoon representation, the catalytic domain in grey, lectin in orange).

As expected, the interaction between TcTS and aminophosphonate compound **48** is the

same as that seen in the crystal structure after soaking with inhibitor alone. Looking at the conformation in the catalytic site, Tyr119 has rotated to the closed position, and the lactose is found between Tyr119 and Trp312. When the TcTS crystal structure complexed with lactose only (PDB: 1MS9) (Buschiazzo. *et al*., 2002) is superimposed, the position of lactose in crystal structure **15** is highly similar (Figure 5.1 B). The binding position of lactose presents that only glucose part of lactose shows stacking with Tyr119 and Trp312, and the distance from center of cyclic center of glucose to the stacking residues is 5.1 Å to Tyr119 and 4.1 Å to Trp312.

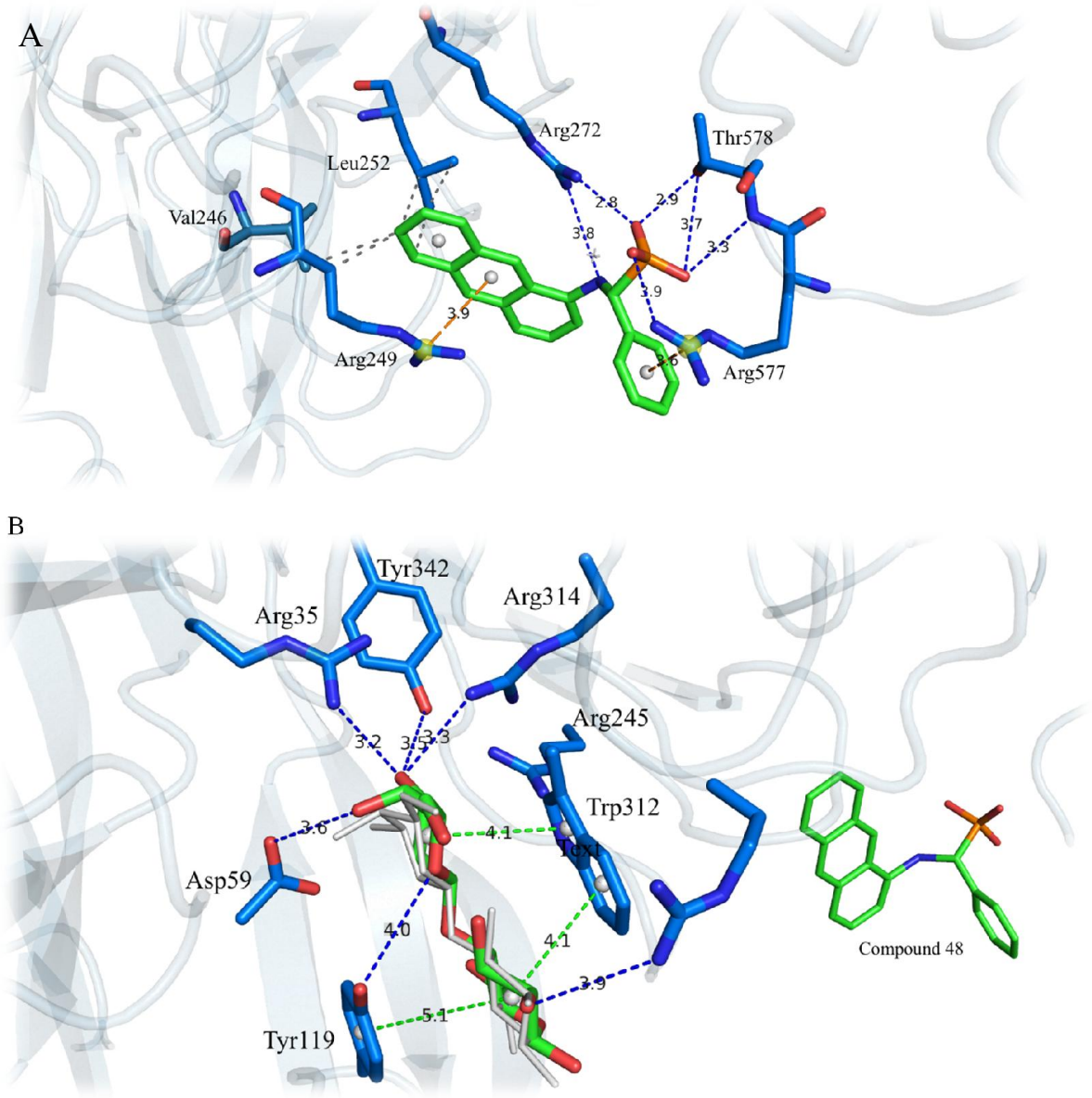


Figure 5.3. A. Detailed interactions of compound 48 with TcTS at the allosteric site. **B.** Details of the interaction of lactose (in green and red stick form) with TcTS at the catalytic site, with

the lactose molecule seen in the crystal structure of TcTS bound with DANA and lactose (PDB: 1MS9) shown in grey bonds following superimposition of the two structures.

Therefore, considering the electronic distribution on either galactose and glucose (Martinez and Iverson, 2012), the lactose molecule may have a weak π-π interaction with both Tyr119 and Trp312. Apart from the weak π-π interaction, the main contribution to the interaction between lactose and TcTS is through hydrogen bonds, several key residues in catalytic site, such as Asp59, Tyr342 and the arginine triad are all interact with lactose. The average distance of the hydrogen bonds between lactose and binding residues is 3.8 Å, indicating that the hydrogen bonding in this interaction is weak (Jeffrey and Jeffrey, 1997). Both weak π-π interaction and weak hydrogen bonding result in an overall weak interaction between lactose and the lactose binding site, this finding could explain the low affinity of lactose to TcTS catalytic site even in the presence of sialic acid. It can also be conjectured that lactose derivatives are difficult to develop as a potential inhibitor due to the weak affinity of TcTS for lactose in the binding site. In contrast, the anthraquinone compound could be a potential inhibitor probing the lactose-binding site of TcTS due to the stronger interactions, since the distance from aromatic centers of anthraquinone to either Tyr119 or Trp312 is less than 4Å, which is shorter than lactose (4 - 5 Å).

Consequently, the crystal structure shows that binding of the α-aminophosphonate compounds induces lactose binding site formation independent of the SA binding site occupation. And stable lactose binding in the lactose-binding site occurs in a crystal soaked with the α-aminophosphonate inhibitor, but not in the absence of inhibitor. This finding suggests that the inhibition of this new class of inhibitors could be achieved through affecting the conformation of the catalytic site. By interaction with α-aminophosphonates, the lactose site has increased stability and this interrupts the release of the hydrolysis product or acceptor substrate from the catalytic site.

5.4. Thermodynamic study of lactose binding to TcTS

The crystal structure suggested that lactose bound in the presence of an α-aminophosphonate. To investigate the interaction of secondary substrate (lactose in this case) in the presence of inhibitor, a thermodynamics study was introduced. ITC was used as the technique to determine the temperature changes of lactose binding to the enzyme-inhibitor complex. In inhibition assays and the ITC study, the inhibition constants IC_{50}, K_i and the binding constant K_d were all detected. Therefore, for this ITC assay, the enzyme was pre-incubated with 0.5 mM of the inhibitor compound **59**, which is more than 10 times the concentration given by the inhibition constant (K_i = 0.014 mM) to insure the interaction between TcTS and inhibitor was saturated before the titration of lactose began.

Site directed mutagenesis of D59A was performed following the standard protocol. The primer pair $TcTS_{D59A}F$ and $TcTS_{D59A}R$ (Appendix 8.5) were purchased from Invitrogen, UK. The PCR reaction was performed using *pfu* DNA polymerase (source). The PCR product was cleaned using a QIAquick PCR purification kit (Thermo Scientific) and digested using Fastdigest DpnI restriction enzyme (Thermo Scientific). The digested PCR product was then transferred to NEB5α and plated onto LB agar plates. Single colonies were picked and amplified by overnight incubation in LB (with appropriate antibiotics), and the plasmids were extracted by GeneJET plasmid miniprep kit (Thermo Scientific). The insert sequence was determined by DNA sequencing (Source bioscience).

Thermodynamic studies of TcTS interacting with lactose were performed in a Nano-ITC (TA instrument, USA). Purified TcTS was concentrated to 0.16 mM, then dialysed against ITC experiment buffer (Appendix 8.2). Sialic acid (10 µl from 50 mM stock in ITC buffer), inhibitor compound **59** (10 µl from 20 mM stock in ITC buffer) or ITC buffer (10 µl) were added into 190 µl dialyzed TcTS and preincubated at 25°C for 30 min. ITC experiments were performed at 25°C, with 300 rpm stirring. ITC data were

collected using NanoRun (TA instrument) and thermodynamic parameters were processed using NanoAnalysis (TA instrument).

A previous study achieved the binding of lactose in the presence of sialic acid using a mutated $TcTS_{D59N}$, which resulted in stable binding of lactose by eliminating catalytic activity (Buschiazzo. *et al*., 2002). In $TcTS_{D59A}$ the TcTS catalytic site has lost a key residue due to the truncation of the Asp59 side chain which plays a vital role in the cleavage of the donor substrate and linkage of the sialyl moiety to the acceptor substrates. Unsurprisingly, this mutation eliminated major hydrolysis activity of TcTS. Surface plasmon resonance (SPR) indicated that the binding of sialic acid or other donor sialyl-oligosaccharides induces the formation of the lactose binding site. The lactose binding experiment was measured by ITC to compare the lactose binding to TcTS in the presence of lactose and sialic acid then lactose would be titrated into TcTS pre-incubated into compound **59,** in order to investigate the interaction between lactose and TcTS-inhibitor complex.

ITC results of mutated $TcTS_{D59N}$ (Figure 5.4) shows that lactose can bind to $TcTS_{D59N}$ in complex with sialic acid and inhibitor compound **59**, with a similar range of heat generation increase (approximately 1.8-0.4μJ/mol) as TcTS (1-0.2 μJ/mol), indicated that stronger interaction occurs in $TcTS_{D59N}$ than TcTS. Crystal structures suggest that lactose binds to the catalytic site mainly by stacking with Tyr119 and Trp312, however, all the ITC experiments using either TcTS or $TcTS_{D59N}$ suggest that the interaction of lactose is driven by enthalpy rather than entropy, meaning that hydrogen bonds and van de Waals interactions play a larger role in the lactose interaction than hydrophobic interactions.

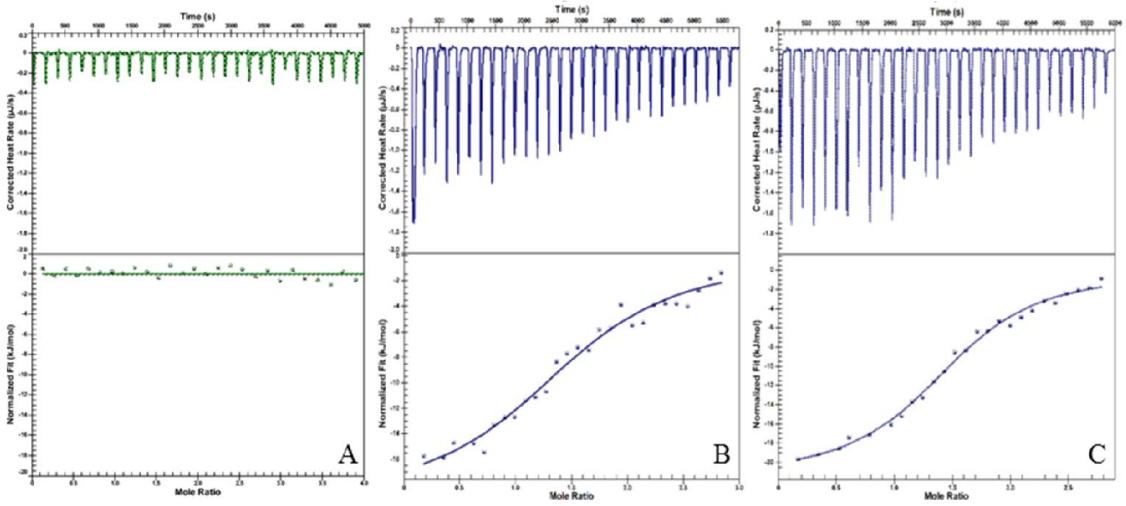


Figure. 5.4. Titration pattern (upper panels) of ITC experiment and fitted curve (lower): **A**. lactose was titrated into $TcTS_{D59N}$, **B**. lactose was titrated into $TcTS_{D59N}$ plus sialic acid; **C**. lactose was titrated into $TcTS_{D59N}$ plus compound **59.**

Wide type TcTS failed to achieve saturation with low concentration of lactose, either in case of only TcTS or with the addition of sialic acid or inhibitor, since lactose bind weakly to TcTS even in the presence of donor substrate (MuNANA, K_m= 0.79 mM) (Wilbrink, M. H. *et al*., 2014), and an ITC assay needs high substrate concentration and K_d ratio (c value) to achieve a proper situation curve (100>c>10) (Turnbull and Daranas, 2003). However, when the concentration of TcTS was increased to 0.25mM, and was then titrated by 50mM lactose, it presented a significant saturation process (figure 5.5), whilst the lactose titrated into native TcTS showed a much lower level of heat generation, suggesting that without the presence of inhibitor, lactose tends not to interact with TcTS.

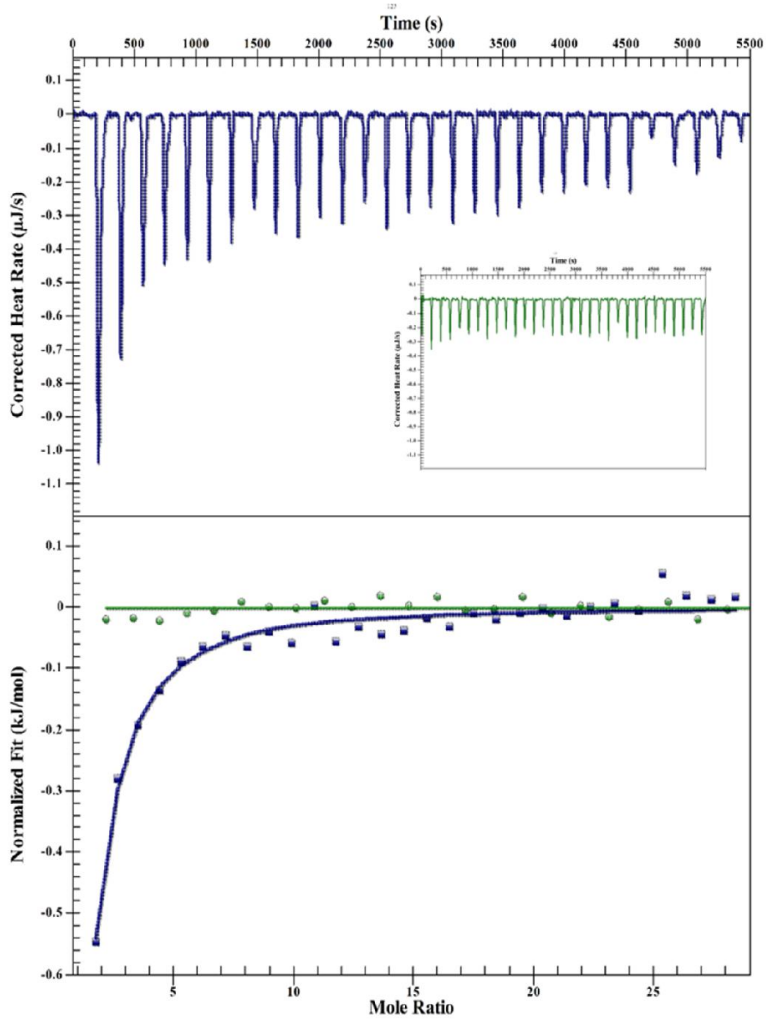


Figure. 5.5. Titration pattern of ITC experiment (upper panel), and fitted curve (lower): Blue: 50 mM lactose titrated into 0.15 mM TcTS preincubated with 2 mM compound **59**. Green 50 mM lactose titrated into 0.15 mM TcTS.

However, the reaction parameters between lactose and TcTS are unable to be determined by ITC because the heat generated by the transfer sialic acid activity might disrupt the signal generated by lactose binding. Additionally, the c value ($c=MK_dn$, where M is starting concentration of protein, K_d is the binding constant and n is the stoichiometry parameter) (Duff, *et al.*, 2011) of the wild type TcTS titration by lactose is lower than 0.5, which will strongly impact the veracity of stoichiometry and entropy changes. The ITC gave a binding constant (K_d) of lactose binding in the presence of inhibitor is 0.46mM, which is similar to the kinetic constant (K_m) of transfer-lactose activity. These results indicate that lactose can interact with TcTS.

5.5. Conclusion

Lactose plays a crucial role in TcTS catalytic activity as an acceptor substrate, thus it is important to evaluate TcTS activity and inhibition in the presence of lactose. In kinetics studies, addition of lactose did not change the overall catalytic efficiency but influenced the kinetics parameters such as K_m and k_{cat}. The changes of kinetics parameter induced by lactose were not affected by α-aminophosphonate inhibitors, illustrating that this type of inhibitor is neither competitive against donor structure nor acceptor substrate in the interaction with TcTS, supporting the crystallographic data that it binds to a novel allosteric site but not in the catalytic pocket.

In the crystal structure soaked with both inhibitor and lactose, the hypothesis of lactose binding site formation has been proved and supported by lactose binding determination in both structural and thermodynamic studies. With induction of lactose-binding site formation, the K_m of lactose binding is unchanged compared to the lactose-binding site formation induced by DANA. Thus, as observed from the crystal structures, the formation of the lactose binding site, which is triggered by either sialic acid or allosteric inhibitors would have similar effect on lactose binding site formation. The presence of α-aminophosphonate did not change the K_m between donor substrate, in this case MuNANA, and the TcTS catalytic site but did reduce the turnover number (k_{cat}) of the hydrolysis activity, with or without lactose. It illustrates that with α-aminophosphonate bound, donor substrates can still interact with the SA binding site, and the most probable inhibitory mechanism is to abolish the catalytic process or prevent substrate release.

This structure characteristic is also proved by thermodynamic studies. In the presence of an α-aminophosphonate compound, lactose shows a significant interaction with the TcTS protein and this interaction tends to be saturated with increasing lactose concentration, suggesting that the binding between lactose and TcTS-inhibitor complex is stable. The ITC results not only present the lactose-TcTS interaction based on

inhibitor binding, but also suggest that the lactose binding is stable, which may explain why catalytic activity will be blocked.

Chapter 6. Conclusions and future Prospects

6.1. Future development of a potential TcTS inhibitor based on the inter-domain allosteric site

The discovery of a TcTS inhibitor has been a great challenge for rational drug design due to its complex catalytic site structure and mechanisms. During the last decade, numerous strategies have been applied to the design of potential TcTS inhibitors, including modification of natural substrates, and high-throughput screening of the databases of known compounds. Unfortunately, unlike other sialidases which can be inhibited by substrate analogues, TcTS cannot be inhibited strongly by substrate-related compounds, especially SA analogues. This failing can be attributed to the unique structure of the TcTS catalytic site which has an extra hydrophobic "gate" (Tyr119 and Trp312), which might affect the recognition and interaction with inhibitor molecules. In addition, the dual-substrate catalytic mechanism is another possible reason to explain the weak inhibition of SA-related inhibitors. Due to TcTS's sialylation activity, many SA analogues, such as DANA, can be used as a donor substrate which will be released from the catalytic site through attack by either water or an acceptor molecule (Buschiazzo. *et al*., 2002). Interestingly, after broad range screening among different small molecule databanks, the most potent inhibitors targeting TcTS were both compounds which are not related to substrates. And numbers of compounds inhibit TcTS activity in a non-competitive manner, which suggests the existance of an extra binding site apart from the catalytic pocket.

This study describes α-aminophosphonate derivatives as a new type of TcTS inhibitor, which were designed by mimicries and simplification of the natural substrates of TcTS. Nevertheless, the inhibition mechanism assays went contrary to our initial plans and suggested non-competitive inhibition. And further kinetics results, including inhibitor removal and time-dependent experiments, supported this result that the inhibition triggered by α-aminophosphonate corresponds was both non-competitive and a non-covalent inhibition. Structural studies through crystal soaking illustrated the non-

competitive and non-covalent binding and indicate a novel allosteric site located approximately 30 Å from the catalytic pocket (Figure 7.1). The electrostatic characteristics of this allosteric site show that this binding site between the two domains of TcTS is mainly positively-charged, which is adequate for the binding of phosphonic acid compounds. None of the catalytic residues are involved in the allosteric binding site, indicating that the inhibition is by an indirect mechanism. And no major conformation change occurs, the interaction at the allosteric site only induces rearrangement of several residues like a closing door.

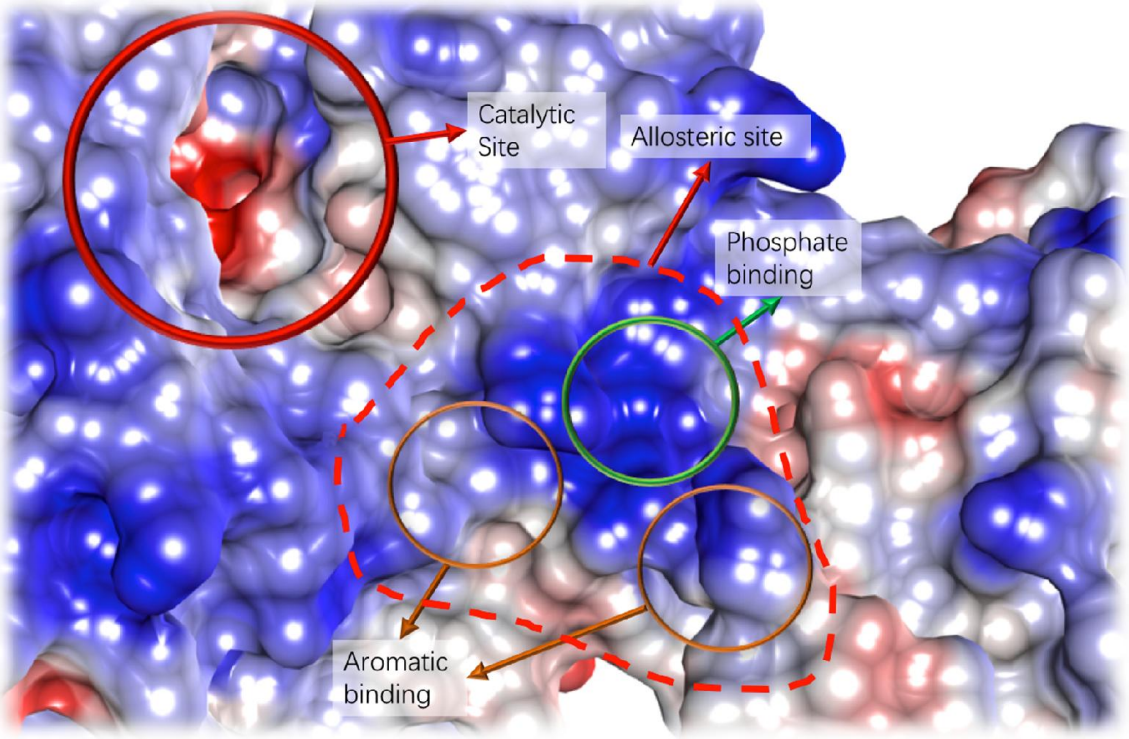


Figure 6.1. The catalytic site and allosteric binding site of TcTS, presented in a surface representation with electrostatic colorings, the positively charged residues are in blue and negatively charged residues are in red. Catalytic site is highlighted by a solid red circle. The allosteric binding site is circled by a red dashed line, in which the phosphonate binding area is highlighted by a green circle and aromatic binding areas by brown circles.

The allosteric binding has three contributing parts (Figure 6.1, Table 6.1), the multiple-rings aromatic group stacking (Arg249, Leu252), benzene ring interaction (Arg577), and phosphonate binding (Arg272, Arg577, Thr578) areas. Amongst all the previous

structural studies including those in this thesis about ligand binding to TcTS, phosphonate compounds are the only molecules that can be identified in this allosteric site, suggesting that the phosphate group is the core of the allosteric interaction. Moreover, the geometry of the phosphate moiety indicates that it possibly fits better with a tetrahedral acidic structure such as phosphate or sulfonate groups (Figure 6.2). In further potential inhibitor design and screening, this hypothesis could be investigated with different substitutions for this moiety.

Table 6.1 Polar and hydrophobic interactions between TcTS and α-aminophosphonates

Residue	Domain	Inhibitor	Inhibitor fragment	Interaction type	Distance (Å)
AM moiety substitutes					
Arg249	Catalytic domain	48, 59, 61	Anthracene	Cation-π interaction	3.3
		62, 63	Anthraquinone	Hydrogen bond	3.4
Leu252		48, 59, 61	Anthracene	Hydrophobic interaction	3.5-4.0
		62, 63	Anthraquinone		
Phosphate group					
Arg272	Catalytic domain	*All	Phosphate	Salt bridge	3.0
Arg577	Lectin-like domain			Salt bridge	2.7-3.0
Thr578	Lectin-like domain			Hydrogen bond	2.5-3.0
AD moiety substitutes					
Arg577	Lectin-like domain	*All	Phenyl	Cation-π interaction	3.0-3.5
		59, 63	Acetamido O	Hydrogen bond	2.8-3.2
		61	Hydroxyl O	Salt bridge	3.8

*All: All inhibitors were studied with XRC (compounds 48, 59, 61, 62, 63)

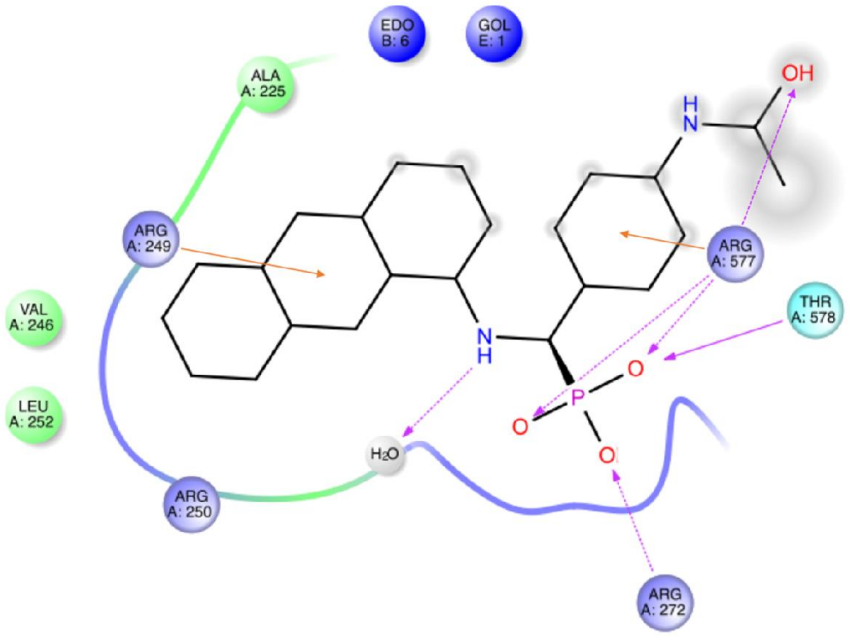


Figure 6.2. 2D view of the interaction between compound 59 and TcTS at the allosteric binding site, hydrogen bond is presented as purple lines, salt bridges as purple dashed lines, and the cation-π interaction is shown as orange lines. Generated by LeView (Caboche, 2013).

Aminosulfonate (**64**) is able to be synthesized as a replacement of phosphonate to form the corresponding sulfonate compounds (Mauro, 2010). Based on the phenyl-phenyl compound, the synthesis of a phenyl-naphthalene compound (**65**) was attempted by an undergraduate project student, but it failed to purify by column chromatography. Nevertheless, the impure sample still gave an IC_{50} of 0.06 mM (data not shown), which is similar to the phenyl-anthracene aminophosphonate compound (**48,** IC_{50} = 0.045 mM). This attempt suggests that aminosulfonate compounds could also be a potential type of inhibitor against TcTS. It can be seen that the nitrogen at the amino group only interacts with solvent but not with any residue of TcTS in the allosteric binding site, indicating that the amino group might not contribute to the ligand affinity and inhibition process. Therefore, replacing the amino-nitrogen by a carbon ion could be beneficial to the chemical stability of the compounds. Looking at phosphonate compounds, phenylphosphonic acid (**66**) has been reported as a human prostatic acid phosphatase inhibitor (Schwender *et al*., 1996), so further modifications can be achieved through a similar synthesis strategy. An α-substituted phenyl sulfonic acid (**67**) has also already

been reported which can be used as the basic skeleton for further modifications (Enders *et al*., 2002).

Based on structural studies and previous research, the phosphate group could be considered as a possible reason for the non-catalytic-pocket binding since all phosphonate inhibitors against TcTS are non-competitive inhibitors. Another strategy in molecular mimicry could be replacement of the phosphate group by a carboxyl group which would maintains the negatively charged substitutent seen in SA. Both the amino acid derivative (**68**) and a compound that replaced the nitrogen with a carbon (**69**) have been reported to be synthesized (Juhl *et al*., 2017; Sathe *et al*., 2013).

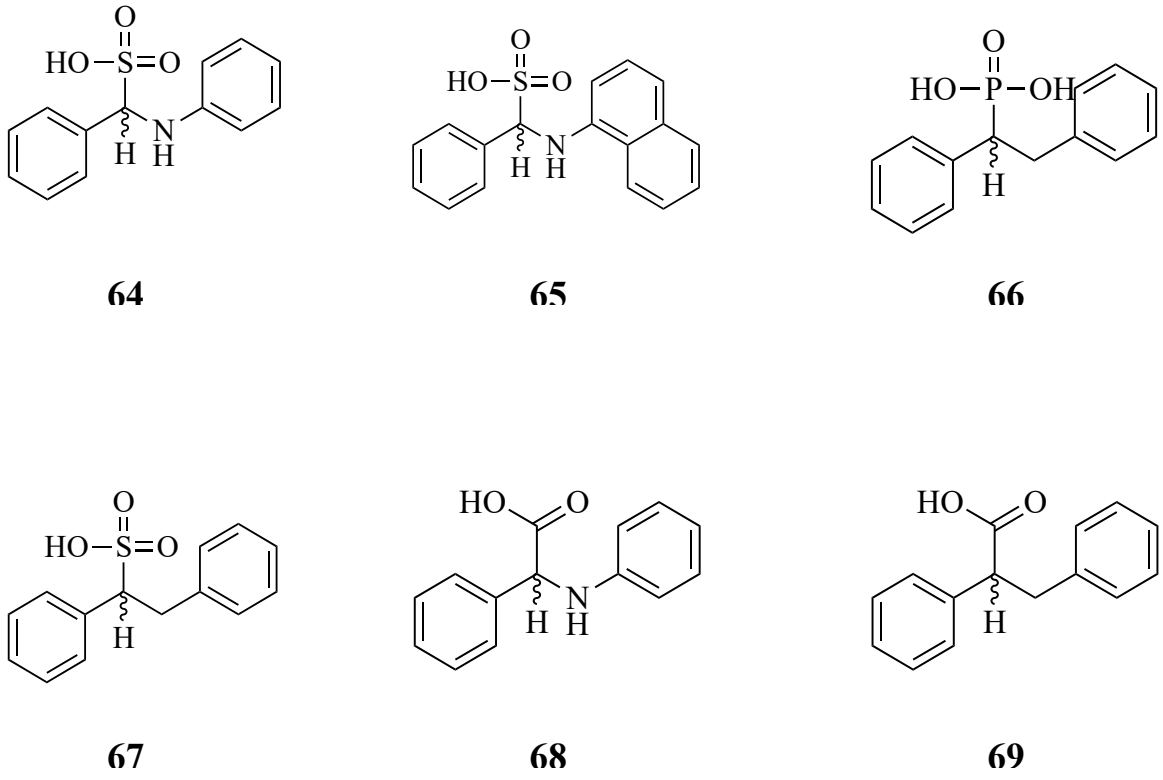


Figure 6.3. Predicted structures of potential TcTS inhibitors based on interaction at the allosteric site

The mechanism of the allosteric inhibition by α-aminophosphonates can be attributed to the formation of the lactose-binding site and induction of lactose binding without the presence of SA substrates. However, the detailed molecular mechanisms that trigger conformation change of the catalytic site indirectly need further investigation. Molecular dynamics (MD) simulation will be a powerful tool to probe the effect on key residues at the catalytic site after interacting with α-aminophosphonates. Further biochemical and biophysical techniques such as mutagenesis, kinetics studies, structural studies, etc. would favor the interpreting of the molecular mechanisms of this allosteric inhibition accomplished with computer-assisted approaches.

6.2. Future inhibitor development based on catalytic site

The catalytic site is the most important site for TcTS activity, as well as for potential inhibitor design. In our XRC study of the known inhibitor Rhein, to our surprise we found that the inhibitor molecule interacts with the lactose binding site, a part of the catalytic site. This finding suggests that lactose-site binding might not affect the interaction between TcTS and its donor substrate. The hydrophobicity distribution of the TcTS catalytic site (Figure 6.4) shows that the areas on both the top and the bottom of the SA binding pocket are hydrophilic, which favors SA interaction. In the closed conformation, the hydrophilic area above the SA binding pocket is reduced and covered behind the “closed door” formed by Tyr119 and Trp312, Therefore, the hydrophilic area below the SA pocket could be important for SA binding. For example, the mutagenesis Asp247 significantly reduced the hydrolysis activity (Oliveira *et al*., 2014).

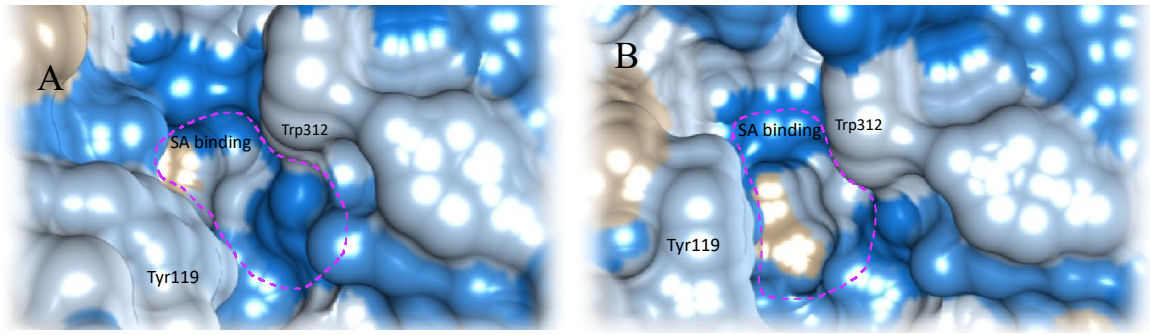
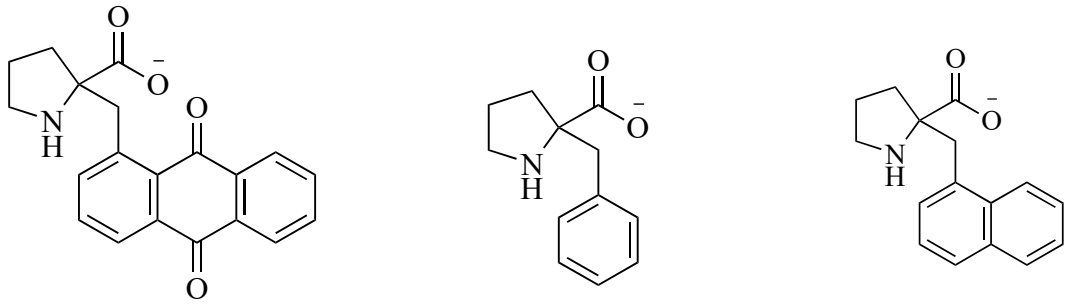


Figure 6.4. The surfaces of the open (A) and closed (B) conformation of the TcTS catalytic pocket, colourd by hydrophobicity (image generated by Chimera). Blue is hydrophilic, white is hydrophobic, and tan is maximum hydrophobic. The lactose binding site is formed by Tyr119 and Trp312, the SA binding site is circled by purple dashed line.

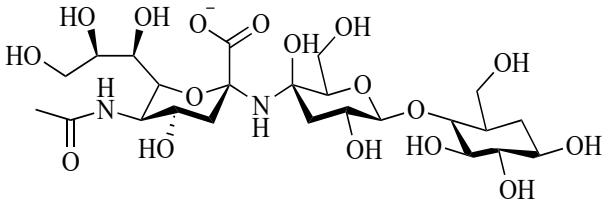
Additionally, the soaked crystal structures show that both naphthalene and anthracene do not interact with the catalytic site, and that anthraquinone and lactitol and other lactose derivatives, are lactose site binders, which can be applied to future drug discovery as useful fragments. In terms of the SA binding site, in all soaked crystal structures, it appears that the phenyl-phosphonate cannot bind to the catalytic site like SA and its derivatives. But fragment-based drug design by Telford indicated that proline could be a potential SA-site binder (Telford, 2014). Therefore, future structural based drug design could be measured based on these findings. For example, the compound that combines proline and anthraquinone (**70**) could be a potential inhibitor which binds to both the SA pocket and lactose binding site, the synthesis could be based on that of the known compounds 2-(phenylmethyl)-L-Proline (**71**) (Sutar and Joshi, 2013) and 2-(1-naphthalenylmethyl)-L-Proline (**72**) (Ilisz *et al*., 2014) which are also potential TcTS inhibitor.



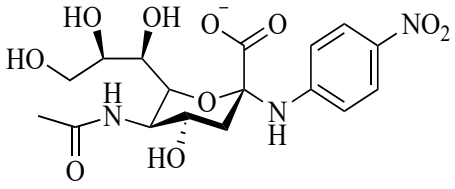
70 **71** **72**

Figure 6.5 Structure of Proline-based predicted TcTS inhibitors.

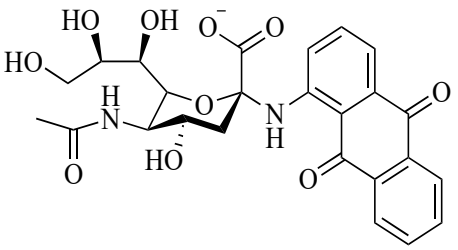
In the interaction between α-aminophosphonate compounds and the TcTS inter-domain allosteric site, the replacement of the glycosidic bond oxygen by a nitrogen did not utilise its function and did not even have any interaction in the allosteric binding. However, in structure-based design based on catalytic site binding fragments, the nitrogen linkage between SA-mimic moiety and lactose-mimic moiety can still be expected to interact with Asp59 through acidic-basic interactions. SA and SA derivatives are not good inhibitors of TcTS, but SA-1-N- compounds might have a better inhibition effect since it would also interact with the lactose-binding site based on SA substrate recognition. For example, with simple alteration of the glycosidic bond into an amino bond on SAL (**73**) it is possible to achieve catalytic site binding without hydrolysis activity, inducing TcTS inactivation. Aromatic groups, for instance, 4-nitrophenyl (**74**, Vibrio cholera neuraminidase inhibitor) (Khorlin *et al*., 1970) and anthraquinone (**75**), can be attached to SA by an N-link for a π-π interaction with the lactose binding site.



73



74



75

Figure 6.6 Structure of predicted TcTS inhibitors based on SA.

6.3. Future directions for the investigation of TcTS catalytic mechanisms

Through the investigation of TcTS-inhibitor complex structure and inhibitory mechanisms, useful information would be provided for further research into the inhibition, catalytic mechanisms, and even further applications of the biosynthesis activity of TcTS. Firstly, TcTS is currently only a potential therapeutic target against Chagas disease because of the lack of strong and specific inhibitor. α-Aminophosphonates are potential inhibitors but still need improved inhibition effect and specificity. *In vivo* pharmacokinetics study and cytotoxicity studies should be measured especially as these are non-natural compounds. Fortunately, the crystal structures of TcTS-inhibitor complex have been solved making future rational development clearer.

After intensive research for more than a decade, the catalytic process of TcTS is basically understood. However, there are still several unsolved questions regarding detailed mechanisms and biological roles. One of the key arguments about the catalytic mechanism is whether the two-step reaction follows a classical ping-pong mechanism or a ternary-complex mechanism. In the classical ping-pong mechanism, the binding of the acceptor substrate in the lactose binding site occurs after the release of the hydrolysis product (Damager *et al*., 2008). However, after the release of the hydrolysis product glycans, the TcTS-SA complex could be attacked by solvent rather than accepter glycans, resulting in a low efficiency of SA transfer activity (Oliveira *et al*., 2014). Therefore, a ternary complex of TcTS, donor substrates and acceptor substrates could prevent miss-attack by solvents. In our crystal structure studies, anthraquinone derivatives, a type of non-competitive inhibitor of TcTS, was identified in the lactose binding site, illustrating the possibility of the ternary complex formation.

Another question for the sequential catalytic process is that it could lead to the re-sialylation of donor glycans, which will also result in low efficiency. From enzyme kinetics, it is reported that TcTS shows selectivity for the acceptor, for example, using MuNANA as the SA donor, TcTS has higher affinity towards 3'-galactosyl-lactose than lactose (Wilbrink, Maarten H *et al*., 2014). However, in the case of TcTS activity *in vivo*, when host sialyl-glycans are structurally similar to *T. cruzi* mucin glycans, how TcTS prevents transfer inaccurately to host glycans and hydrolysis of *T. cruzi* sialyl mucin glycans is still unclear. The channel between catalytic site and the inter-domain allosteric site is a potential binding site for longer glycan chains since multiple polar small molecules such as glycerol or PEG derivatives bind (Figure 6.7), indicating that this area may have some function in substrate recognition. Thus, the future studies about TcTS catalytic details could focus on two major aspects, the first being the intermediate ternary complex, and the second the substrate recognition. A better understanding of the TcTS mechanism could favour future drug candidate development, in order to develop an effective chemical therapy against Chagas disease. And the interpretation of knowledge of substrate recognition could be applied to bio-synthesis of sialyl-glycan

products using TcTS as a catalyst.

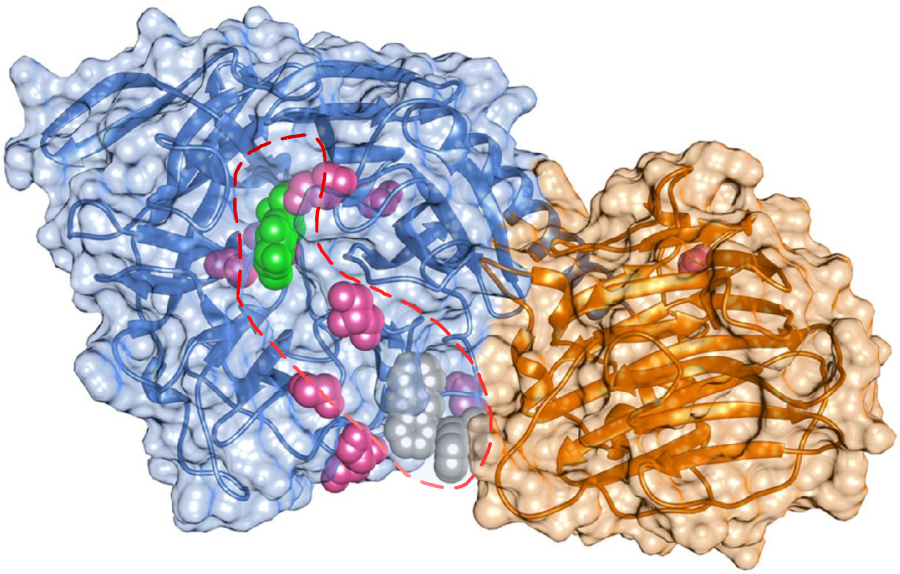


Figure 6.7. Overview of crystal structure of TcTS soaked with lactose and compound 48. Catalytic domain residues are shown in blue and lectin-like domain in orange. Compound 48 is in grey space-filling representation, lactose is in green and glycerol and ethylene glycol are in pink.

Chapter 7. Experimental methods

/

7.1. Materials

7.1.1. Chemicals, reagents and solvents.

Table 7.1. List of chemicals, solvent and their suppliers.

Chemical	**Supplier**
1 kb plus DNA ladder	Invitrogen
1-Butanol	Fisher scientific
1-Naphthaldehyde	Alfa-asear
1-Naphthylamine	Sigma-Aldrich
2-aminoanthracene	Sigma-Aldrich
2-aminoanthraquinone	Alfa-asear
2-O- (4-methylumbelliferyl) -a-D-N-acetylneuraminic acid	carbosynth
2-Propanol	Sigma-Aldrich
3,5-Bis(trifluoromethyl)benzaldehyde	Sigma-Aldrich
36.5% formaldehyde solution	Sigma-Aldrich
4-(Trifluoromethyl)benzaldehyde	Sigma-Aldrich
4-Acetamidobenzaldehyde	Sigma-Aldrich
4-methylumbelliferone sodium salt	Sigma-Aldrich

Chemical	**Supplier**
4-Nitrobenzaldehyde	Sigma-Aldrich
4-Nitrobenzamide	Sigma-Aldrich
6X DNA loading dye	Thermo scientific
Acrylamide 40% solution	Fisher scientific
Agarose, electrophoresis grade	Melford
Ampicillin	Sigma-Aldrich
BamH I restriction enzyme	New England Biolabs
Benzaldehyde	Sigma-Aldrich
Benzaldehyde	Sigma-Aldrich
Phenylamine	Sigma-Aldrich
Bovine Serum Albumin (BSA)	Sigma-Aldrich
Brilliant Blue R-250	Fisher scientific
Bromophenol Blue	Sigma-Aldrich
Bromotrimethylsilane purum	Sigma-Aldrich
Calcium chloride	Sigma-Aldrich
Deoxyribo nucleotide PCR grade (dNTPs)	Thermo scientific

Chemical	**Supplier**
Dimethyl phosphite	Sigma-Aldrich
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich
EDTA- disodium salt	Sigma-Aldrich
Ethidium bromide	Sigma-Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
Fast digest DpnI restriction enzyme	Thermo scientific
Glycerol	Fisher scientific
Glycine	Fisher scientific
HEPES	Sigma-Aldrich
Kanamycin sulfate	Sigma-Aldrich
Lactose anhydrous	carbosynth
LB-Agar miller	Fisher scientific
Magnesium chloride	Fluca
Magnesium sulfate	Sigma-Aldrich
N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid(DANA)	Sigma-Aldrich
N-Acetylneuraminic acid (Sialic acid, NANA)	Sigma-Aldrich

Chemical	**Supplier**
Nickel sulfate	Sigma-Aldrich
PEG 4000	Sigma-Aldrich
Pfu DNA Polymerase	Agilent
Phosphate buffer saline tablets	Oxoid
Phusion DNA Polymerases	Thermo scientific
Polyethylene glycol (PEG) 3350	Sigma-Aldrich
Potassium sulfate	Sigma-Aldrich
RNAse A solution	Promega
Sodium acetate	Sigma-Aldrich
Sodium chloride	Fisher scientific
Sodium dodecyl sulphate (SDS)	Sigma-Aldrich
Sodium hydrogen carbonate	Fisher scientific
T4 DNA ligase	New England Biolabs
T4 DNA ligase 10X buffer	New England Biolabs
Tetramethylethylenediamine (TEMED)	Sigma-Aldrich
Thermoprime plus DNA polymerase	Thermo scientific

Chemical	**Supplier**
Tris-HCl	Sigma-Aldrich
Triton X-100	BDH laboratory
Trizma base	Fisher scientific
Tryptone	BD
Xhol I restriction enzyme	New England Biolabs
Yeast extract	BD
Solvent	**Supplier**
Acetone HPLC grade	Fisher
Dichloromethane	Fisher
Ethanol	Fisher
Ethyl acetate	Fisher
Ethyl ether	Fisher
Methanol	Fisher
Petroleum ether	Fisher
Toluene	Fisher
Water, HPLC grade	Fisher

7.1.2. Plasmids, cell strains and media

7.1.2.1. Expression plasmids

The plasmid TcTS-pTrcHisA encoding recombinant TcTS truncated to contain only the catalytic and lectin-like domains with 7 mutations (Asn58Phe, Ser495Lys, Val496Gly, Glu520Lys, Asp593Gly, Ile597Asp, and His599Arg) on the surface of protein (performed to produce a crystallizable TcTS construct (Buschiazzo. *et al*., 2002)) was provided as a gift from the Prof. P. Alzari group (Pasteur Institute, Paris, France). None of these mutations directly involve in catalytic progress, although Asn58Phe is close to the catalytic site, and the mutations will not affect TcTS activity. The pTrcHisA (ampicillin resistance, Appendix 8.1) expression plasmid is designed for effective expression of recombinant protein controlled by the *pTrc (lac)* promotor (Li *et al*., 1984), and contains a 6- histidine tag on the N-terminus...

Plasmid SpCBM-pHisTEV was provided as a gift from Prof. G. Taylor group (University of St. Andrews, UK). pHisTEV vector (kanamycin resistance, Appendix 8.1), which is modified from the pET21d expression vector, is used for expression of N-terminal His-tagged protein with the tag cleavable with TEV protease.

7.1.2.2. Cell strains

The *E. coli* strains XL-Gold Ultra and XL-Blue (Agilent) chemically competent cells were used for TcTS molecular cloning experiments. For mutagenesis experiments and cloning of SpCBMcat recombinant protein, *E. coli* strain NEB5α (NEB, UK) was applied. The *E. coli* strain BL21 (DE3) (NEB) chemically competent cells were used for expression of recombinant protein.

7.1.2.3. Media

E. coli cells were cultured in autoclaved sterilized *luria-bertani* broth (LB, recipe given in Appendix table 8.2.), which is a nutritionally-rich medium. All components were weighed and dissolved into 900ml distilled water then autoclaved and stored at room temperature or 4°C for longer storage. Antibiotic stock solutions (100mg/ml ampicillin or 50mg/ml kanamycin) were added to the medium at a dilution of 1/1000 (v/v) prior to the cell culture. Transformed competent cells were spread onto LB agar plates containing 2% agar powder in LB media. The LB agar solution was autoclaved and allow to cool down to 45°C before adding the antibiotic solutions, the media was poured into 90mm Petri dishes with 10ml per dish. The LB agar plates were sealed and stored at 4°C.

TB based auto-induction medium (TBAIM) was used to express proteins cloned into pTrc vectors where protein expression is controlled by *trc* promotor. Like LB medium, TB medium also contain tryptone, and yeast extract, but it also has additional 0.5% of glycerol, and 10x TB salt which contains 0.17M KH_2PO_4 and 0.7M K_2HPO_4, which would prepared and autoclaved separately (Appendix 8.2). After cooling down, 100ml of 10x TB salt solution was added to 900ml TB medium and the appropriate antibiotics added before cell culture.

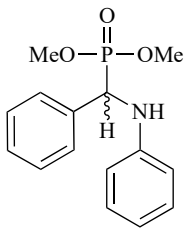
After transformation, *E. coli* cells were recovered in super optimal medium with catabolite repression (SOC, Appendix 8.2). SOC media is a rich media with multiple salts that favor the recovery of bacteria culture after transformation (Hanahan, 1983).

7.2. Experimental: Chemistry.

7.2.1. General

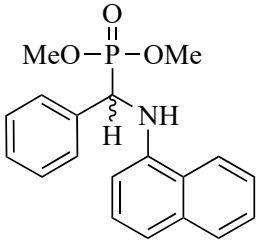
Chemical reagents were purchased from Sigma-Aldrich unless specifically stated. Anhydrous solvents were purchased from Sigma-Aldrich and used without further purification. All other solvents were purchased from Fisher Scientific. Analytical thin layer chromatography (TLC) was carried out on Merck aluminium backed TLC plates silica gel 60 F254 (0.25 mm thickness), viewed using UV light of wavelength 254 nm or stained with potassium permanganate solution or *para*-anisaldehyde stain. Silica gel chromatography was performed on silica gel 60 Å (200-400 mesh) from Sigma-Aldrich. Standard work-up is referring to diluting the reaction residue in appropriate organic solvent (mentioned), washing with an aqueous saturated sodium bicarbonate solution, followed by washing with brine. The organic phases were then combined and dried over anhydrous magnesium sulphate, filtered and the filtrate concentrated *in vacuo*. Preparative reverse phase chromatography (C-18) was performed using a VersaFlash hand held column (23 x 110 mm) from Supelco. Melting points were obtained using a Thermo Fisher IA9000 digital melting point apparatus. ^{1}H, ^{13}C and ^{31}P NMR spectra were recorded using Bruker Advance III (400 and 500 MHz) spectrometers with acquisition frequencies of 400 or 500 MHz for ^{1}H; 101 or 125 MHz for ^{13}C; and 162 MHz for ^{31}P. Deuterated solvents were purchased from Cambridge Isotope Laboratories. The NMR chemical shifts δ are recorded in parts per million (ppm) with reference to tetramethylsilane for ^{1}H and ^{13}C and phosphoric acid for ^{31}P NMR spectroscopy. The multiplicities are assigned as a singlet (s), doublet (d), triplet (t), quartet (q), doublet of doublets (dd), doublet of doublet of doublets (ddd), doublet of triplets (dt), triplet of doublets (td), broad (br) and multiplet (m). High resolution mass spectrometry was performed using a BrukerMicrOTOF electrospray ionisation mass spectrometer.

7.2.2. Synthesis of protected aminophosphonate



S41. Dimethyl (phenyl (phenylamino) methyl) phosphonate

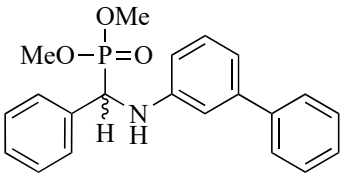
A mixture of benzaldehyde (0.48 mL, 4.7 mmol) and analine (0.45 mL, 4.7 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (5 mL, 10.0 equiv. v/v) and the resultant solution subjected to microwave heating (90 W) for 5 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (35 % EtOAc/petroleum ether) to give the protected aminophosphonate **S41** as a white solid (1.02 g, 75 %). Spectroscopic data (^{1}H and ^{13}C NMR) was consistent with that reported in the literature (Tajbakhsh *et al*., 2008). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.48 – 7.40 (m, 2H, Ar), 7.35 – 7.26 (m, 2H, Ar), 7.26 – 7.19 (m, 1H, Ar), 7.12 – 7.00 (m, 2H, Ar), 6.66 (t, *J* = 7.3Hz, 1H, Ar), 6.60 – 6.53 (m, 2H, Ar), 4.81-4.70(m, 2H, NH, P-CH), 3.71 (d, *J* = 10.6 Hz, 3H, OMe), 3.43 (d, *J* = 10.6 Hz, 3H, OMe).^{13}C NMR (100.6 MHz, $CDCl_3$): δ 146.16, 129.13, 128.65, 127.98, 127.76, 118.48, 113.85, 55.68 (d, *J* = 151.3 Hz, P-CH), 53.75, 53.69. ^{31}P NMR (161.93 MHz, $CDCl_3$): δ 24.96. HRMS (ESI): Calculated for $C_{15}H_{19}NO_3P$ $(M+H)^+$: m/z = 292.1024, found m/z = 292.1066. Calculated for $C_{15}H_{18}NO_3PNa$ $(M+Na)^+$: m/z= 314.0844., found m/z = 314.0907.



S42. Dimethyl ((naphthalen-1-ylamino) (phenyl) methyl) phosphonate

A mixture of benzaldehyde (0.48 ml, 4.7 mmol, 1.0 equiv.) and 1-aminonapthalene

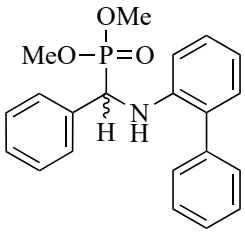
(0.14 g, 4.7 mmol, 1 equiv.) were dissolved in dimethyl phosphite (5 mL, 10.0 equiv. v/v) and the resultant solution subjected to microwave heating (90 W) for 5 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (60 % EtOAc/petrol) to give the protected aminophosphonate **S42** as a white solid (1.36 g, 85 %). Spectroscopic data (^{1}H and ^{13}C NMR) was consistent with that reported in the literature (Mu *et al*., 2006). ^{1}H NMR (400 MHz, CDCl3): δ 8.06-8.03 (m, 1H, Ar), 7.79 (d, *J* = 8.0Hz, 1H, Ar), 7.56 – 7.45 (m, 4H, Ar), 7.38 – 7.22 (m, 4H, Ar), 7.19-7.16 (m, 1H, Ar), 6.49 – 6.37 (m, 1H, Ar), 5.49 (s, 1H, NH), 3.80 (d, *J* = 10.6 Hz, 3H, OMe), 3.53 (d, *J* = 10.6 Hz, 3H, OMe). ^{13}C NMR (101 MHz, CDCl3) δ 141.05, 135.25, 134.20, 128.72, 128.62, 128.07, 127.58, 126.13, 125.85, 125.11, 123.88, 119.97, 118.69, 106.54, 55.87 (d, *J* = 151.0 Hz, P-CH), 53.85, 58.81. 31P NMR (162 MHz, CDCl3): δ 24.94. HRMS (ESI) calculated for $C_{19}H_{21}NO_3P$ $[M+H]^+$: 342.1225. Found: 342.1233.



S44. Dimethyl (([1,1'-biphenyl]-3-ylamino)(phenyl)methyl)phosphonate

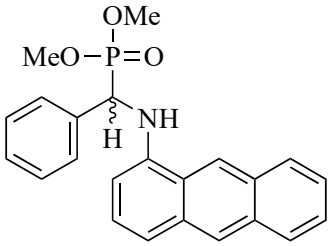
A mixture of benzaldehyde (0.48, 4.7 mmol, 1.0 equiv.) and 3-Aminobiphenyl (0.79 g, 4.7 mmol, 1.0 equiv.) were dissolved in dimethylphosphite (5 mL, 10.0 equiv. v/v) and the resultant solution subjected to microwave heating (90 W) for 7 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (32 % EtOAc/petrol) to give the protected aminophosphonate (**S44**) as a white solid (1.34 g, 84 %). Spectroscopic data: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.50 – 7.43 (m, 2H, Ar), 7.43 – 7.37 (m, 2H, Ar), 7.36 – 7.27 (m, 4H, Ar), 7.27 – 7.24 (m, 1H, Ar), 7.23 – 7.18 (m, 1H, Ar), 7.11 (t, *J* = 7.8 Hz, 1H, Ar), 6.89-6.86 (m, 1H, Ar), 6.79 (t, *J* = 2.0 Hz, 1H, Ar), 6.54-6.51 (m, 1H, Ar), 5.22 (s, 1H, NH), 4.81 (br. d, *J* = 25.0

Hz, 1H, P-CH), 3.72 (d, *J* = 10.7 Hz, 3H, OMe), 3.43 (d, *J* = 10.6 Hz, 3H, OMe). ^{13}C NMR (101 MHz, $CDCl_3$) δ 146.42, 135.56, 129.49, 128.70, 128.52, 128.03, 127.79, 127.73, 127.11, 126.96, 117.57, 112.81, 112.67, 55.71 (d, *J* = 151.4 Hz, P-CH), 53.76, 53.69. ^{31}P NMR (162 MHz, $CDCl_3$): δ 24.88. HRMS (ESI) calculated for $C_{21}H_{24}N_2O_4P$ $[M+H]^+$: 368.1414. Found: 368.1415. Calculated for $C_{21}H_{23}N_2O_4PNa$ $[M+H]^+$: 390.1235. Found: 390.1230.



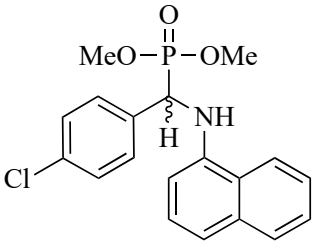
S45. Dimethyl (([1,1'-biphenyl]-2-ylamino) (phenyl) methyl) phosphonate

A mixture of benzaldehyde (0.48 ml, 4.7 mmol, 1.0 equiv.) and 2-aminobiphenyl (0.79 g, 4.7 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (5 mL, 10.0 equiv. v/v) and the resultant solution subjected to microwave heating (90 W) for 5 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (30 % EtOAc/petrol) to give the protected aminophosphonate (**S45**) as a white solid (1.3 g, 76.4 %). Spectroscopic data: ^{1}H NMR (400 MHz, CDCl3): δ 7.53 – 7.44 (m, 4H, Ar), 7.40-7.38 (m, 3H, Ar), 7.37 – 7.31 (m, 2H, Ar), 7.31 – 7.27 (m, 1H, Ar), 7.14 – 7.09 (m, 2H, Ar), 6.79 (t, *J* = 7.4Hz, 1H, Ar), 6.64 – 6.46 (m, 1H, Ar), 4.99 (s, 1H NH), 4.80 (br. d, *J* = 24.2 Hz, 1H, P-CH), 3.62 (d, *J* = 10.6 Hz, 3H, OMe), 3.47 (d, *J* = 10.6 Hz, 3H, OMe). ^{13}C NMR (100.6 MHz, $CDCl_3$): δ 143.14, 138.88, 135.61, 130.25, 129.18, 128.92, 128.78, 128.62, 127.96, 127.62, 127.50, 118.28, 111.89, 56.13 (d, *J* = 150.8 Hz, P-CH), 53.69, 53.52. ^{31}P NMR (162 MHz, $CDCl_3$): δ 24.51. HRMS (ESI) calculated for $C_{21}H_{24}N_2O_4P$ $[M+H]^+$: 367.1337. Found: 368.1414. Calculated for $C_{21}H_{23}N_2O_4PNa$ $[M+H]^+$: 390.1235. Found: 390.1247.



S48. Dimethyl ((anthracen-1-ylamino) (phenyl) methyl) phosphonate

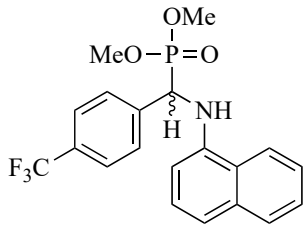
A mixture of benzaldehyde (0.48 g, 4.7 mmol, 1.0 equiv.) and 1-aminoanthracene (0.91 g, 4.7 mmol 1.0 equiv.) were dissolved in dimethyl phosphite (5 mL, 10.0 equiv. v/v) and the resultant solution subjected to microwave heating (90 W) for 9 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (80 % EtOAc/petrol) to give the protected aminophosphonate (**S48**) as a brown solid (1.1 g, 87 %). Spectrocopic data: ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.59 (s, 1H, Ar), 8.36 (s, 1H, Ar), 8.16 – 8.08 (m, 1H, Ar), 8.02 – 7.94 (m, 1H, Ar), 7.61 – 7.53 (m, 2H, Ar), 7.53 – 7.44 (m, 2H, Ar), 7.44 – 7.24 (m, 4H, Ar), 7.20 – 7.10 (m, 1H, Ar), 6.40 – 6.33 (m, 1H, Ar), 5.70 (s, 1H, NH), 5.06 (d, *J* = 23.9 Hz, 1H, P-CH) 3.85 (d, *J* = 10.7 Hz, 3H, OMe), 3.54 (d, *J* = 10.5 Hz, 3H, OMe). ^{13}C NMR (100.6MHz, $CDCl_3$): δ 140.63, 135.17, 132.34, 131.57, 131.05, 128.75, 128.44, 128.10, 127.70, 127.63, 127.58, 126.71, 125.61, 125.27, 123.81, 118.89, 118.63, 104.51, 55.83 (d, *J* = 151.4 Hz, P-CH), 53.97, 53.91. ^{31}P NMR (162 MHz, $CDCl_3$): δ 24.94. HRMS (ESI) calculated for $C_{23}H_{23}NO_3P$ $[M+H]^+$: 392.1432. Found: 392.1415.



S50. Dimethyl ((4-chlorophenyl) (naphthalen-1-ylamino) methyl) phosphonate

A mixture of 4-Chlorobenzaldehyde (0.5g, 4.7mmol, 1.0 equiv.) and 1-naphthylamine (0.672g, 4.7mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (5 ml, 10.0 equiv. v/v) and the resultant solution subjected to microwave heating (90 W) for 5 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by

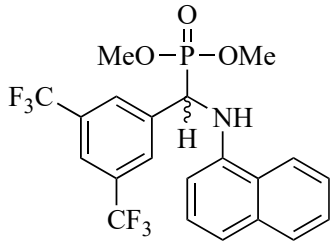
column chromatography (70% EtOAc/PetEther) to give compound **S50** at white solid. Yield: 0.98g (87%). Spectroscopic data: ^{1}H-NMR (400MHz, $CDCl_3$): δ 8.02-7.99 (m, 1H, H-Ar), 7.81-7.77 (m, 1H, H-Ar), 7.55-7.44 (m, 4H, H-Ar), 7.33-7.30(m, 2H, H-Ar), 7.27 (d, 1H, *J*=8.4Hz, H-Ar), 7.19-7.15 (t, 1H, *J*=16Hz, H-Ar), 6.36 (d, 1H, *J*=7.2Hz, H-Ar), 5.42 (s, 1H, N-H), 4.98(d, 1H, *J*=24Hz, P-CH),3.82 (d, 3H,, *J*=10.8Hz, OMe), 3.62 (d, 3H, *J*=10.8Hz, OMe). ^{13}C NMR (100.6MHz, $CDCl_3$): δ 140.83, 140.69, 134.21, 133.93, 133.89, 128.96, 128.93, 128.85, 128.68, 126.04, 125.96, 125.84, 123.85, 119.87, 119.02, 106.59, 56.09 (d, *J* = 150.3 Hz, P-CH), 54.59, 54.08. ^{31}P-NMR (161.93MHz, $CDCl_3$): δ 24.10. HRMS: Calculated for [$C_{19}H_{21}ClNO_3P$]: m/z=375.0869, found [M+H]$^{+}$ =376.0850. Calculated for [$C_{19}H_{20}ClNO_3PNa$] m/z= 397.0688, found [M+Na]$^{+}$ =398.0685.



S51. Dimethyl((naphthalen-1-ylamino) (4-(trifluoro-methyl) phenyl) methyl) phosphonate

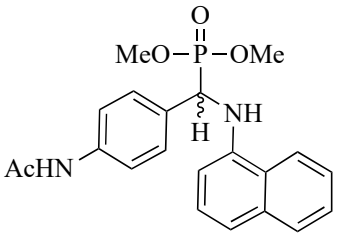
A mixture of 4-(Trifluoromethyl) benzaldehyde (0.5g, 2.87 mmol, 1.0 equiv.) and 1-naphthylamine (0.41g, 2.8 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (5 ml, 10.0 equiv.) and the resultant solution subjected to microwave heating (90 W) for 7 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (70% EtOAc/PetEther) to give compound **S51** at white solid. Yield: 0.65g (55.5%). Spectroscopic data: ^{1}H-NMR (400MHz, $CDCl_3$): δ 8.23 (d, 2H, *J*=6.6Hz, H-Ar), 7.96 (s, 1H, H-Ar), 7.82 (d, 2H, *J*=6.3Hz, H-Ar,), 7.71-7.52 (m, 4H, H-Ar), 7.24-7.15 (m, 2H, H-Ar), 6.34 (q, 1H, H-Ar, *J*=8.2Hz), 5.53 (s, 1H, N-H), 5.08(d, 1H, P-CH, *J*=21.6Hz),3.85 (d, 3H, OMe, *J*=10.8Hz), 3.69 (d, 3H, OMe, *J*=10.2Hz). ^{13}C NMR (100.6MHz, $CDCl_3$): δ 141.85, 140.66, 140.16, 139.28, 135.37, 132.46, 128.88, 128.48, 127.92, 126.43, 126.08, 125.65, 125.27, 124.36, 122.36,

120.07, 117.04, 58.31 (d, *J* = 153.3 Hz, P-CH), 53.97, 53.67, ^{31}P-NMR (161.93MHz, $CDCl_3$): δ 24.60. HRMS: Calculated for [$C_{20}H_{20}F_3NO_3P$]: m/z=409.1124, [$C_{20}H_{19}F_3NO_3Na$] m/z= 431.0952, found [M+Na]$^+$ =432.0942.



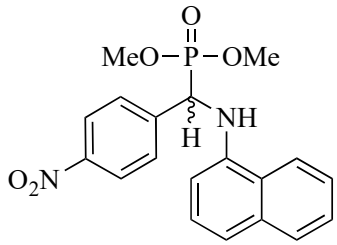
S52.Dimethyl ((3,5-bis(trifluoromethyl) phenyl) (naphthalen-1-ylamino) methyl) phos-phonate

A mixture of 3,5-Bis(trifluoromethyl)benzaldehyde (0.5g, 2.06 mmol, 1.0 equiv.) and 1-naphthylamine (0.295g, 2.06 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (5ml, 10 equiv.) and the resultant solution subjected to microwave heating (90 W) for 7 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (80% EtOAc/PetEther) to give compound **S52** at white solid. Yield: 0.82g (83.6%). Spectroscopic data: ^{1}H-NMR (400MHz, $CDCl_3$): δ 8.04 (t, 3H, *J*=17.6Hz, H-Ar), 7.83 (t, 2H, *J*=8.2Hz, H-Ar), 7.58-7.50 (m, 2H, H-Ar), 7.32 (d, 1H, H-Ar, *J*=8.4Hz), 7.21 (t, 1H, *J*=16Hz), 6.29 (d, 1H, *J*=5.6Hz, H-Ar), 5.48 (s, 1H, NH), 5.08(d, 1H, *J*=24.4Hz, P-CH), 3.85 (d, 3H, OMe, *J*=10.8Hz), 3.69 (d, 3H, *J*=10.8Hz, OMe). ^{13}C NMR (100.6MHz, $CDCl_3$): δ 142.65, 140.15, 140.06, 137.18, 133.27, 131.56, 128.95, 128.58 (d, C-Ar, J=2.4Hz), 127.87, 126.23, 126.17, 125.37, 125.07, 124.76, 122.36, 121.83, 120.05, 116.04, 56.35 (d, *J* = 153.3 Hz, P-CH), 54.07, 53.97, ^{31}P-NMR (161.93MHz, $CDCl_3$): δ 24.23. HRMS: Calculated for [$C_{21}H_{19}F_6NO_3P$]: m/z=477.1007, found [M+H]$^+$ =478.1019.14. Calculated for [$C_{21}H_{18}F_6NO_3PNa$] m/z=499.0826., [M+Na]$^+$ =500.0837.



S53. Dimethyl ((4-acetamidophenyl) (naphthalen-1ylamino) methyl) phosphonate

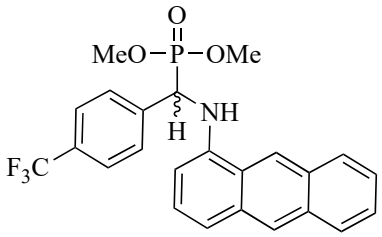
A mixture of 4-Acetamidobenzaldehyde (0.5g, 3.06 mmol, 1.0 equiv.) and 1-naphthylamine (0.437g, 3.06 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (5ml, 10 equiv.) and the resultant solution subjected to microwave heating (90 W) for 8 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (70% EtOAc/petrol) to give compound **S53** at white solid. Yield: 1.31g (72%). Spectroscopic data: 1H-NMR (400MHz, CDCl3): δ 8.00 (d, 1H, *J*=6.4Hz, H-Ar), 7.83-7.71 (m, 1H, H-Ar), 7.53-7.71 (m, 6H, H-Ar),7.25 (d, 1H, *J*=8Hz, H-Ar), 7.18-7.14 (t, 1H, *J*=16Hz, H-Ar), 6.41 (d, 1H, *J*=6Hz, H-Ar), 5.39 (s, 1H, N-H), 4.97(d, 1H, *J*=20Hz, P-CH),3.81 (d, 3H, *J*=10.4Hz, OMe), 3.57 (d, 3H, *J*=9.6Hz, OMe), 2.15(s, 3H, Ac-NH). ^{13}C NMR (100.6MHz, $CDCl_3$): δ 163.63, 141.15, 141.06, 138.28, 134.37, 130.65, 128.85, 128.38, 128.07, 126.33, 126.07, 125.37, 124.06, 123.33, 120.05, 119.04, 55.55 (d, *J* = 150.5 Hz, P-CH) 54.23, 54.07, 29.86, 24.63. ^{31}P-NMR (161.93MHz, CDCl3): δ 24.71. HRMS: Calculated for [C21H24N2O4P]: m/z=398.14, found $[M+H]^+$=399.14. Calculated for [C21H22N2O4PNa] m/z= 420.12, found $[M+Na]^+$ =421.13.



S54. Dimethyl ((naphthalen-1-ylamino) (4-nitrophenyl) methyl) phosphonate

A mixture of 4-Nitrobenzaldehyde (0.5g, 3.3 mmol) and 1-naphthylamine (0.473g, 3.3 mmol) were dissolved in dimethyl phosphite (5ml, 10 equiv.) and the resultant solution subjected to microwave heating (90 W) for 6 minutes. T The reaction mixture then

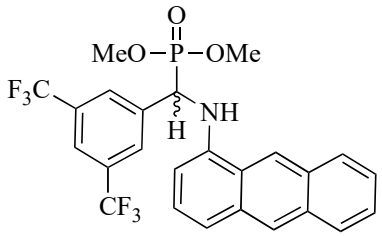
subjected to a standard work-up (DCM) and purified by column chromatography (85% EtOAc/PetEther) to give compound **S54** at yellow solid. Yield: 0.92g (72%). Spectroscopic data: ^{1}H-NMR (400MHz, $CDCl_3$): δ 8.20 (d, 2H, *J*=6.8Hz, H-Ar), 8.01 (d, 1H, *J*=6.4Hz, H-Ar), 7.85 (d, 2H, *J*=5.8Hz, H-Ar), 7.71-7.68 （t, 2H, *J*=8.8, H-Ar）, 7.57-7.49 (m, 2H, H-Ar), 7.27 (d, 1H, *J*=7.2Hz, H-Ar), 7.17-7.13（t, 1H, *J*=16, H-Ar）, 6.27 (d, 1H, *J*=7.6Hz, H-Ar), 5.46 (s, 1H, NH), 5.10(d, 1H, *J*=24.8Hz, P-CH),3.92 (d, 3H, *J*=10.2Hz, OMe), 3.68 (d, 3H, *J*=9.9Hz, OMe). ^{13}C NMR (100.6MHz, $CDCl_3$): δ 143.63, 143.28, 140.06, 134.32, 128.77, 128.43, 128.38, 126.16, 125.91, 125.49, 123.92, 123.88, 123.84, 119.75, 119.53, 106.60, 55.41 (d, *J* = 150.5 Hz, P-CH) 54.21, 53.93. ^{31}P-NMR (161.93MHz, $CDCl_3$): δ 23.20. HRMS: Calculated for [$C_{19}H_{20}N_2O_5P$]: m/z=386.1099, Found $[M+H]^+$=387.1109. Calculated for [$C_{19}H_{19}N_2O_5PNa$] m/z= 408.0925. $[M+Na]^+$ =409.0925.



S57. Dimethyl ((anthracen-1-ylamino) (4- (trifluoromethyl) phenyl) methyl) phosphonate

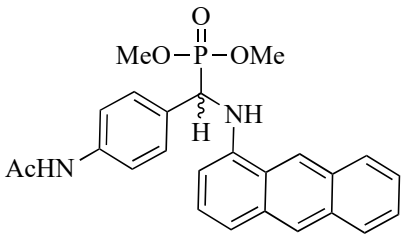
A mixture of 4-(Trifluoromethyl) benzaldehyde (0.5g, 2.8 mmol, 1.0 equiv.) and 1-naphthylamine (0.54g, 2.8 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (6 ml, 10.0 equiv.) and the resultant solution subjected to microwave heating (90 W) for 9 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (90% EtOAc/PetEther) to give compound **S57** at brown solid. Yield: 0.32g (21%). Spectroscopic data: ^{1}H-NMR (400MHz, $CDCl_3$): δ 8.37 (s, 1H, H-Ar), 8.37 (s, 1H, H-Ar), 8.14 (d, 2H, *J*=8.8Hz, H-Ar), 8.03-7.88 (m, 3H, H-Ar), 7.68 (s, 1H, H-Ar), 7.55-7.51 (m, 3H, H-Ar), 7.18 (d, 1H, *J*=7.6Hz, H-Ar), 6.34 (d, 1H, *J*=8.8Hz, H-Ar), 5.83 (s, 1H, N-H), 5.17(d, 1H, *J*=24.1Hz, P-CH), 3.88 (d, 3H, *J*=9.8Hz,

OMe), 3.31 (d, 3H, *J*=10.6Hz, OMe). ^{13}C NMR (100.6MHz, $CDCl_3$): δ 142.38, 139.64, 137.53, 133.46, 132.04, 131.94, 131.62, 128.80, 128.04, 127.34, 126.80, 126.37, 125.48, 125.15, 125.04, 123.47, 122.23, 117.73, 116.82, 106.89, 56.53 (d, *J* = 150.3 Hz, P-CH), 54.37, 54.25. ^{31}P-NMR (161.93MHz, $CDCl_3$): δ 24.34. HRMS: Calculated for [$C_{24}H_{21}F_3NO_3P$]: m/z=459.1289, Found $[M+H]^+$ =460.1275. Calculated for [$C_{23}H_{21}NO_3PNa$] m/z= 481.1131. $[M+Na]^+$ =482.1120.



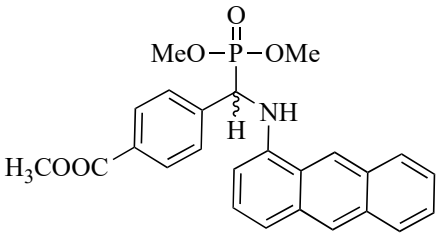
S58. Dimethyl ((anthracen-1-ylamino) (3,5-bis (trifluoromethyl) phenyl) methyl) phosphonate

A mixture of 3,5-Bis(trifluoromethyl)benzaldehyde (0.5g, 2.06 mmol, 1.0 equiv.) and 1-naphthylamine (0.397g, 2.06 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (6.5 ml, 10.0 equiv.) and the resultant solution subjected to microwave heating (90 W) for 7 minutes. The reaction mixture then subjected to a standard work-up (DCM) and purified by column chromatography (90% EtOAc/PetEther) to give compound **S58** at brown solid. Yield: 0.83g (75.6%). Spectroscopic data: ^{1}H-NMR (400MHz, $CDCl_3$): δ 8.57 (s, 1H, H-Ar), 8.39 (s, 1H, H-Ar), 8.14 (t, 2H, *J*=9.2Hz, H-Ar), 8.04 (q, 2H, *J*=20.8H,z H-Ar), 7.84 (s, 1H, H-Ar), 7.52-7.47 (m, 3H, H-Ar), 7.20 (t, 1H, H-Ar, *J*=16Hz), 6.22 (d, 1H, *J*=8.6Hz, H-Ar), 5.72 (q, 1H, *J*=16.8Hz, NH), 5.17(d, 1H, *J*=24Hz, P-CH) 3.89 (d, 3H, *J*=10.6Hz, OMe), 3.35 (d, 3H, *J*=10.8Hz, OMe). ^{13}C NMR (100.6MHz, $CDCl_3$): δ 140.18, 140.04, 138.93, 132.26, 131.94, 131.74, 131.19, 128.40, 127.79, 127.64, 127.00, 125.88, 125.58, 125.25, 124.44, 123.60, 122.219, 121.73, 119.93, 118.52, 104.59, 55.82 (d, *J* = 150.5 Hz, P-CH), 54.93, 54.35. ^{31}P-NMR (161.93MHz, $CDCl_3$): δ 22.54. HRMS: Calculated for [$C_{25}H_{20}F_6NO_3P$]: m/z=527.1163, found $[M+H]^+$ =528.1150. Calculated for [$C_{25}H_{19}F_6NO_3Na$] m/z= 549.0934., $[M+Na]^+$ =550.0977.



S59. Dimethyl ((4-acetamidophenyl) (anthracen-1-ylamino) methyl) phosphonate

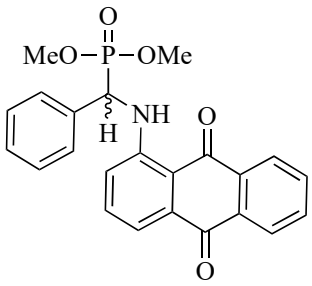
A mixture of 4-acetamidobenzaldehyde (0.5 g, 3.07 mmol, 1.0 equiv.) and 1-aminoanthracene (0.59 g, 3.07 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (5 mL, 10.0 equiv. w/v) and the resultant solution subjected to microwave heating (90 W) for 8 minutes. The reaction mixture then subjected to a standard work-up and purified by column chromatography (2 % MeOH/EtOAc) to give the protected aminophosphonate (**S59**) as a brown solid (0.36 g, 43 %). Spectrocopic data: ^{1}H-NMR (400MHz, CDCl3): δ 8.55 (s, 1H, Ar), 8.35 (s, 1H, Ar), 8.15 – 8.06 (m, 1H, Ar), 8.03 – 7.92 (m, 1H, Ar), 7.48 (d, *J* = 9.6 Hz, 6H, Ar), 7.41 (d, *J* = 8.5 Hz, 1H, Ar), 7.18 – 7.12 (m, 1H, Ar), 6.33 (d, *J* = 7.3 Hz, 1H, Ar), 5.61 (s, 1H, NH), 5.02 (d, *J* = 23.7 Hz, 1H, P-CH), 3.85 (d, *J* = 10.6 Hz, 3H, OMe), 3.57 (d, *J* = 10.6 Hz, 3H, OMe). ^{13}C NMR (101MHz, $CDCl_3$): δ 168.24, 140.48, 137.94, 132.30, 131.56, 131.03, 130.61, 128.41, 128.23, 127.71, 126.75, 125.61, 125.31, 123.80, 119.95, 119.01, 118.54, 104.69, 55.31 (d, *J* = 152.3 Hz), 54.01, 53.95, 24.51. ^{31}P-NMR (161.93MHz, CDCl3): δ 24.75. HRMS (ESI): Calculated for $C_{25}H_{26}N_2O_4P$ $[M+H]^+$: 449.1552. Found $[M+H]^+$: 449.1713. Calculated for $C_{25}H_{25}N_2O_4PNa$ $[M+Na]^+$: 471.1371. Found $[M+Na]^+$: 471.1448.



S60. methyl 4-((anthracen-1-ylamino) (dimethoxyphosphoryl) methyl) benzoate

A mixture of benzaldehyde (0.5 g, 3.05 mmol, 1.0 equiv.) and 1-aminoanthracene (0.59 g, 3.05 mmol, 1.0 equiv.) were dissolved in dimethyl phosphite (3 mL, 10.0 equiv. w/v) and the resultant solution subjected to microwave heating (90 W) for 7 minutes. The

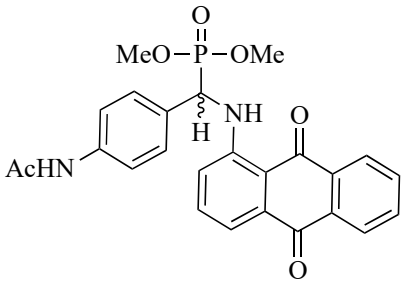
reaction mixture then subjected to a standard work-up and purified by column chromatography (90 % EtOAc/petrol) to give the protected aminophosphonate (**S8**) as a brown solid (0.99 g, 72 %). Spectroscopic data: ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.58 (d, *J* = 1.1 Hz, 1H, Ar), 8.36 (s, 1H, Ar), 8.15 – 8.07 (m, 1H, Ar), 8.07 – 8.00 (m, 2H, Ar), 8.00 – 7.94 (m, 1H, Ar), 7.68 – 7.61 (m, 2H, Ar), 7.53 – 7.45 (m, 2H, Ar), 7.42 (d, *J* = 8.7 Hz, 1H, Ar), 7.18 – 7.10 (m, 1H, Ar), 6.35 – 6.16 (m, 1H, Ar), 5.70 (s, 1H, NH), 5.11 (d, *J* = 24.4 Hz, 1H, P-CH), 3.89 (s, 3H, COOMe), 3.86 (d, *J* = 10.7 Hz, 3H, OMe), 3.60 (d, *J* = 10.7 Hz, 3H, OMe). ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.60, 140.74, 140.44, 132.30, 131.64, 131.11, 130.01, 129.98, 128.41, 127.73, 127.62, 126.81, 125.70, 125.40, 123.76, 119.28, 118.56, 104.64, 55.95 (d, *J* = 149.8 Hz), 54.08, 53.99, 52.05. ^{31}P NMR (162 MHz, $CDCl_3$) δ 24.10. HRMS (ESI): Calculated for $C_{25}H_{25}NO_5P$ $[M+H]^+$: 450.1392. Found $[M+H]^+$: 450.1436. Calculated for $C_{25}H_{24}NO_5PNa$ $[M+Na]^+$: 472.1212. Found $[M+Na]^+$: 471.1288.



S62. Dimethyl (((9,10-dioxo-9,10-dihydroanthracen-1-yl) amino) (phenyl) methyl) phosphonate

A mixture of benzaldehyde (0.48 ml, 4.7 mmol, 1 equiv.) and 1-aminoanthraquinone (1.05 g, 4.7 mmol) were dissolved in dimethyl phosphite (5 mL, 10 equiv. w/v) and the resultant solution subjected to microwave heating (90 W) for 8 minutes. The reaction mixture then subjected to a standard work-up and purified by column chromatography (85% EtOAc/petrol) to give the protected aminophosphonate (**S62**) as a red solid (1.35 g, 89%). Spectroscopic data: ^{1}H NMR (400 MHz, $CDCl_3$) δ 10.73 (t, *J* = 8.8 Hz, 1H, NH), 8.43 – 8.31 (m, 1H, Ar), 8.29 – 8.18 (m, 1H, Ar), 7.80-7.70 (m, 2H, Ar), 7.68 – 7.57 (m, 1H, Ar), 7.57 – 7.48 (m, 2H, Ar), 7.48 – 7.28 (m, 4H, Ar), 6.93 – 6.84 (m, 1H,

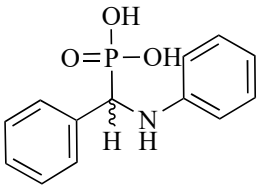
Ar), 5.04 (d, *J* = 31.1 Hz, 1H,P-CH), 3.76 (d, *J* = 10.7 Hz, 3H, OMe), 3.69 (d, *J* = 10.7 Hz, 3H, OMe). ^{13}C NMR (101 MHz, $CDCl_3$) δ 1 185.63, 183.41, 150.09, 135.23, 134.68, 134.64, 133.98, 133.24, 132.91, 128.86, 128.32, 127.69, 127.64, 127.02, 126.70, 118.36, 116.86, 114.61, 55.03 (d, *J* = 152.6 Hz), 54.17, 53.95. ^{31}P NMR (162 MHz, $CDCl_3$) δ 23.12. HRMS (ESI): Calculated for $C_{23}H_{20}NO_5P$ $[M+H]^+$: 479.1294. Found $[M+H]^+$: 479.1379. Calculated for $C_{23}H_{19}NO_5PNa$ $[M+Na]^+$: 500.1113. Found $[M+Na]^+$: 501.1198



S63. Dimethyl (((9,10-dioxo-9,10-dihydroanthracen-1-yl) amino) (phenyl) methyl) phosphonate

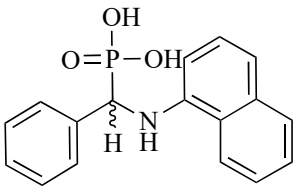
A mixture of 4-acetamidobenzaldehyde (0.5 g, 163.19 mmol, 1 equiv.) and 1-amino-9,10-anthraquinone (0.68 g, 3.06 mmol, 1 equiv.) were dissolved in dimethyl phosphite (5 mL, 10 equiv. w/v) and the resultant solution subjected to microwave heating (90 W) for 5 minutes. The reaction gives the protected aminophosphonate (**S63**) as a red solid which is insoluble in most solvent and only partly soluble in DMSO. The product was sequentially washed by large volume of dichloromethane, petroleum ether, ethyl acetate, methanol and water, then dried and stored at -20°C for next step of reaction. Spectroscopic data: HRMS (ESI): Calculated for $C_{25}H_{23}N_2O_6P$ $[M+H]^+$: . Found $[M+H]^+$: 450.1436. Calculated for $C_{25}H_{22}N_2O_6PNa$ $[M+Na]^+$: 472.1212. Found $[M+Na]^+$: 471.1288.

7.2.3. Synthesis of de-protected aminophosphonic acid



41. (phenyl(phenylamino)methyl) phosphonic acid

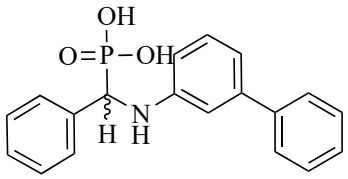
TMSBr (0.22 mL, 1.71 mmol, 5.0 equiv.) was added dropwise to a solution of the aminophosphonate **S41** (100 mg, 0.34 mmol, 1.0 equiv.) in anhydrous DCM (2 mL) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (35 % MeOH/H_2O) to give the phosphonic acid **compound 41** as a white solid (84 mg, 94 %). Spectroscopic data (^{1}H and ^{13}C NMR) was consistent with that reported in the literature (Ryabukhin *et al*., 2014). ^{1}H NMR (400 MHz, CD_3OD): δ 7.54 (d, *J* = 8.0 Hz, 2H, Ar), 7.23 (t, *J* = 7.6 Hz, 2H, Ar), 7.10 (t, *J* = 7.3 Hz, 1H, Ar), 7.03 – 6.89 (m, 2H, Ar), 6.60 – 6.44 (m, 3H, Ar), 4.49 (d, *J* = 21.6 Hz, 1H，P-CH). ^{13}C NMR (100.6 MHz, CD_3OD): δ 150.25, 144.04, 129.62, 129.15, 128.62, 126.46, 117.34, 114.74, 60.64 (d, *J* = 130.3 Hz). ^{31}P NMR (161.93 MHz, CD_3OD): δ 15.46. HRMS (ESI): Calculated for $C_{13}H_{14}NO_3P$ $(M-H)^-$: m/z = 262.0633, found m/z = 262.0633.



42. ((naphthalen-1-ylamino) (phenyl) methyl) phosphonic acid

TMSBr (0.16mL, 0.88 mmol, 3.0 equiv.) was added dropwise to a solution of the aminophosphonate **S42** (100 mg, 0.29 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution

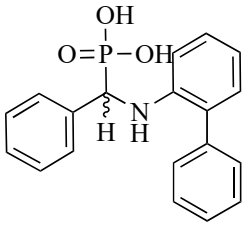
was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (40% MeOH/H_2O) to give the phosphonic acid **compound 42** as a white solid (74.5 mg, 82 %). ^{1}H NMR (400 MHz, CD_3OD): δ 8.37 – 8.29 (m, 1H, Ar), 7.73 – 7.62 (m, 1H, Ar), 7.57 (d, *J* = 8.1 Hz, 2H, Ar), 7.48 – 7.30 (m, 2H, Ar), 7.19 (t, *J* = 7.6 Hz, 2H, Ar), 7.11 – 6.93 (m, 3H, Ar), 6.26 (d, *J* = 7.3 Hz, 1H, Ar), 4.64 (d, *J* = 21.3 Hz, 1H, P-CH). ^{13}C NMR (100.6 MHz, CD_3OD): δ 145.19, 143.55, 135.87, 129.13, 129.07, 128.60, 127.42, 126.46, 126.24, 125.27, 124.96, 122.25, 116.77, 106.31, 60.57 (d, *J* = 130.6 Hz). ^{31}P NMR (162 MHz, CD_3OD): δ 15.74. HRMS (ESI) calculated for $C_{17}H_{15}NO_3P$ $[M-H]^-$: 313.0789. Found: 313.0796.



44. (([1,1'-biphenyl]-3-ylamino)(phenyl)methyl)phosphonic acid

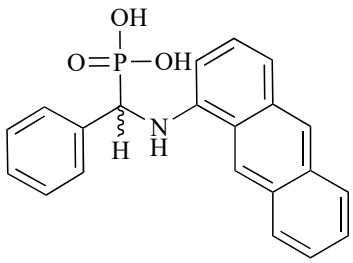
TMSBr (0.16 mL, 1.23 mmol, 4.5 equiv.) was added dropwise to a solution of the aminophosphonate **S44** (100 mg, 0.27 mmol, 1.0 equiv.) in anhydrous DCM (2 mL) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (25 % MeOH/H_2O) to give the phosphonic acid **compound 44** as a white solid (68 mg, 74 %). Spectrocopic data: ^{1}H NMR (400 MHz, Methanol-d_4) δ 7.57 (d, *J* = 7.9 Hz, 2H, Ar), 7.43 – 7.36 (m, 2H, Ar), 7.30 (t, *J* = 7.5 Hz, 2H, Ar), 7.23 (q, *J* = 7.5 Hz, 3H, Ar), 7.09 (t, *J* = 7.3 Hz, 1H, Ar), 7.02 (t, *J* = 7.8 Hz, 1H, Ar), 6.78 (s, 1H, Ar), 6.75 – 6.70 (m, 1H, Ar), 6.53 (d, *J* = 7.9Hz, 1H, Ar), 4.54 (d, *J* = 21.5 Hz, 1H, P-CH). ^{13}C NMR (101 MHz, CD_3OD) δ 150.62, 144.09, 143.43, 142.86, 130.06, 129.53, 129.23, 128.69, 127.86, 126.50, 116.19, 113.90, 113.38, 60.82

(d, *J* = 129.8 Hz). HRMS (ESI) calculated for $C_{19}H_{18}NO_3P$ [M-H]$^-$: 338.0946. Found: 338.0962.



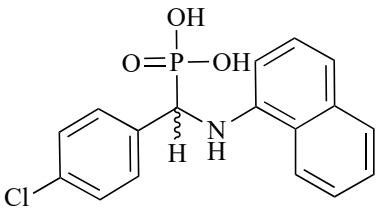
45. (([1,1'-biphenyl]-2-ylamino) (phenyl)methyl)phosphonic acid

TMSBr (0.16 mL, 1.23 mmol, 4.5 equiv.) was added dropwise to a solution of the aminophosphonate **S45** (100 mg, 0.27 mmol, 1.0 equiv.) in anhydrous DCM (2 mL) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (20% MeOH/H_2O) to give the phosphonic acid **compound 45** as a white solid (78.6 mg, 86 %). ^{1}H NMR (400 MHz, CD_3OD): δ 7.61 – 7.54 (m, 2H, Ar), 7.53 – 7.42 (m, 4H, Ar), 7.35 – 7.27 (m, 1H, Ar), 7.22 (t, *J* = 7.6 Hz, 2H, Ar), 7.08 (t, *J* = 7.2 Hz, 1H, Ar), 6.98 – 6.86 (m, 2H, Ar), 6.56 (t, *J* = 7.4Hz, 1H, Ar), 6.46 (d, *J* = 8.4 Hz, 1H, Ar), 4.51 (d, *J* = 21.8 Hz, 1H, P-CH).^{13}C NMR (100.6 MHz, CD_3OD): δ 146.57, 144.08, 141.77, 131.25, 130.51, 129.99, 129.29, 129.05, 128.65, 127.88, 126.49, 117.18, 113.26, 60.63 (d, *J* = 130.9 Hz). ^{31}P NMR (162 MHz, CD_3OD): δ 16.28. HRMS (ESI) calculated for $C_{19}H_{18}NO_3P$ [M-H]$^-$: 338.0946. Found: 338.0960.



48. ((anthracen-1-ylamino) (phenyl)methyl) phosphonic acid

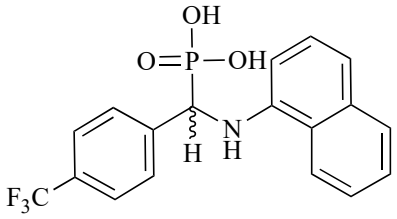
TMSBr (0.289 mL, 2.19 mmol, 8.0 equiv.) was added dropwise to a solution of the aminophosphonate **S48** (100 mg, 0.26 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (60 % MeOH/H_2O) to give the phosphonic acid **compound 48** as a yellow solid (68 mg, 74 %). Spectroscopic data. ^{1}H NMR (400 MHz, CD_3OD): δ 9.00 (s, 1H, Ar), 8.23 (s, 1H, Ar), 8.18 – 8.08 (m, 1H, Ar), 7.98 – 7.87 (m, 1H, Ar), 7.69 – 7.57 (m, 2H, Ar), 7.48 – 7.34 (m, 2H, Ar), 7.25 – 7.12 (m, 3H, Ar), 7.11 – 6.99 (m, 2H, Ar), 6.16 (d, J = 7.4 Hz, 1H, Ar), 4.69 (d, J = 21.0 Hz, 1H, P-CH). ^{13}C NMR (100.6 MHz, CD_3OD): δ 142.94, 140.92, 134.37, 132.87, 132.19, 129.79, 129.16, 128.57, 127.48, 126.57, 126.31, 126.17, 125.51, 125.37, 121.12, 116.50, 113.43, 103.68, 60.76 (d, J = 129.8 Hz, P-C). ^{31}P NMR (162 MHz, CD_3OD): δ 15.57. HRMS (ESI) calculated for $C_{21}H_{17}NO_3P$ $[M-H]^-$: 362.1224. Found 362.1209.



50. ((4-Chlorophenyl) (naphthalen-1-ylamino) methyl) phosphonic acid

TMSBr (0.158 mL, 1.21 mmol, 4.5 equiv.) was added dropwise to a solution of the aminophosphonate **S50** (100 mg, 0.27 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution

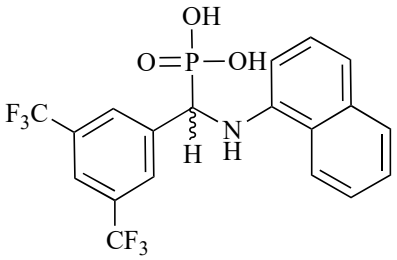
was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (60 % MeOH/H_2O) to give the phosphonic acid **compound 50** as a white solid (78.3mg, 79%). Spectroscopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 8.33 (d, 1H, *J* = 8.6Hz, H-Ar), 7.68-7.66 (m, 1H, H-Ar), 7.55-7.53 (m, 1H, H-Ar), 7.43-7.36 (m, 1H, H-Ar), 7.17 (d, 1H, *J* = 8.4Hz, H-Ar), 7.06-6.99 (m, 2H, H-Ar), 6.2 (q, 1H, *J* = 8.0Hz, H-Ar), 4.79 (d, 1H, *J* = 21.2Hz, P-CH). ^{13}C NMR (100.6 MHz, CD_3OD: δ 146.76, 147.94, 146.89, 144.98, 144.01, 132.43, 129.74, 129.50, 129.15, 126.64, 126.52, 125.37, 122.91, 122.40, 116.59, 106.32, 60.78 (d, *J* = 130.4 Hz, P-C). ^{31}P-NMR (CD_3OD, 161.93 MHz): δ 16.54. HRMS: [$C_{17}H_{15}ClNO_3P$]: m/z = 347.0399, Found $[M-H]^-$ = 356.0414.



51. ((Naphthalen-1-ylamino) (4-(trifluoromethyl)phenyl) methyl) phosphonic acid

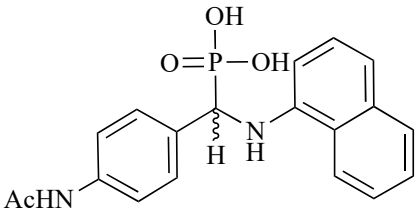
TMSBr (0.182 mL, 1.22 mmol, 5 equiv.) was added dropwise to a solution of the aminophosphonate **S51** (100 mg, 0.244 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (80 % MeOH/H_2O) to give the phosphonic acid **compound 51** as a white solid (62.5 mg, 64.4 %). Spectroscopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 8.35 (d, 1H, *J* = 8.4Hz, H-Ar) 7.65 (d, 1H, *J* = 7.6Hz, H-Ar,), 7.23 (d, 1H, *J* = 7.6Hz, H-Ar), 7.63-7.39 (m, 4H, H-Ar), 7.08-6.95 (m, 3H, H-Ar), 6.38 (d, *J* = 8.2Hz, H-Ar), 4.82 (d, 1H, *J* = 24.1Hz, P-CH). ^{13}C NMR (100.6 MHz, CD_3OD): 149.08, 145.05, 135.919, 129.43, 129.39, 129.17, 128.59, 128.30, 127.63, 127.40, 126.37, 125.38, 125.26, 125.26, 125.08, 124.94, 122.29, 106.15, 61.44 (d, *J* = 129.7 Hz,

P-C). ^{31}P-NMR (161.93 MHz, CD_3OD): δ 17.53. HRMS: [$C_{18}H_{15}F_3NO_3P$]: m/z = 381.0663, Found $[M-H]^-$ = 380.0690.



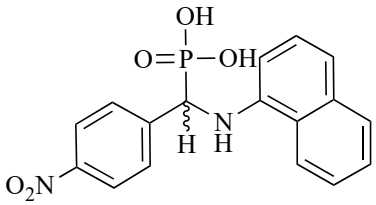
52. ((3,5-bis(trifluoromethyl) phenyl) (naphthalen-1-ylamino) methyl) phosphonic acid

TMSBr (0.092 mL, 0.73 mmol, 3.5 equiv.) was added dropwise to a solution of the aminophosphonate **S52** (100 mg, 0.21 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated in vacuo. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (80 % MeOH/H_2O) to give the phosphonic acid **compound 52** as a white solid (38.3 mg, 39.5 %). Spectroscopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 8.39 (d, 1H, *J* = 8.0Hz, H-Ar), 8.18 (s, 1H, H-Ar), 7.72 (d, 1H, *J* = 8.0Hz, H-Ar), 7.63 (s, 1H, H-Ar), 7.48-7.39 (m, 2H, H-Ar), 7.07 (q, 2H, *J* = 11.2Hz, H-Ar) 6.15 (q, 1H, *J* = 8.4Hz, H-Ar), 4.77 (d, 1H, *J* = 20.0Hz, P-CH), ^{13}C NMR (100.6 MHz, CD_3OD): 148.35, 148.30, 144.79, 135.94, 131.82, 131.47, 131.15, 129.32, 129.20, 127.34, 126.75, 126.74, 126.52, 125.28, 124.09, 122.27, 122.07, 117.54, 105.93, 61.43 (d, *J* = 129.9 Hz, P-C). 31P-NMR (161.93 MHz, CD_3OD): δ 13.53. HRMS: [C19H14F6NO3P]: m/z = 449.0537, Found $[M-H]^-$ = 448.0567.



53. ((4-Acetamidophenyl) (naphthalen-1-ylamino) methyl) phosphonic acid

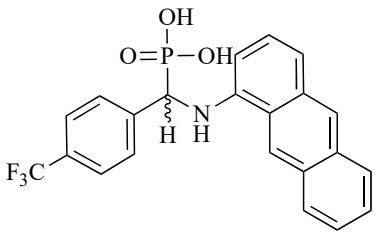
TMSBr (0.08 mL, 0.62 mmol, 2.5 equiv.) was added dropwise to a solution of the aminophosphonate **S53** (100 mg, 0.25 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated in vacuo. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (80 % MeOH/H_2O) to give the phosphonic acid **compound 53** as a white solid (86 mg, 93.5 %). Spectroscopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 8.34 (d, 1H, *J* = 8.4Hz, H-Ar), 7.67 (d, 1H, *J* = 10.4Hz, H-Ar), 7.54-7.51 (m, 2H, H-Ar), 7.41-7.33 (m, 4H, H-Ar), 7.04-6.95 (m, 2H, H-Ar), 6.26 (d, 1H, *J* = 7.6Hz, H-Ar), 4.63 (d, 1H, *J* = 20.8Hz, P-CH), 2.063 (s, 3H, Me). ^{13}C NMR (100.6 MHz, CD_3OD): δ 171.41, 145.16, 139.97, 137.05, 135.76, 129.19, 129.15, 129.04, 127.33, 126.17, 125.16, 122.17, 121.02, 116.70, 106.25, 60.77(d, *J* = 129.6 Hz, P-C), 23.61. ^{31}P-NMR (161.93 MHz, CD_3OD): δ 15.01. HRMS: [$C_{19}H_{19}N_2O_4P$]: m/z = 370.1082, Found $[M-H]^-$ = 369.1005.



54. ((Naphthalen-1-ylamino) (4-nitrophenyl) methyl) phosphonic acid

TMSBr (0.148 mL, 1.125 mmol, 4.5 equiv.) was added dropwise to a solution of the aminophosphonate **S54** (100 mg, 0.25 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated in vacuo. The resultant

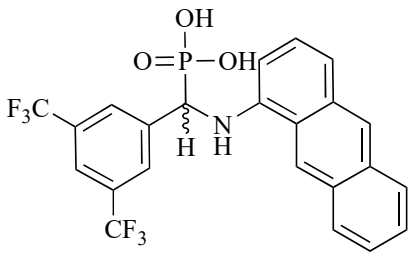
residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (50 % MeOH/H_2O) to give the phosphonic acid **compound 54** as a yellow solid (54.4 mg, 59.3 %). Spectroscopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 8.37 (d, 1H, *J* = 8.4Hz, H-Ar), 8.091 (d, 1H, *J* = 8.8Hz, H-Ar), 7.79 (d, 2H, *J* = 8.4Hz, H-Ar), 7.69 (d, 1H, *J* = 8.0Hz, H-Ar), 7.49-7.38 (m, 2H, H-Ar), 7.032 (d, 2H, *J* = 4.0Hz, H-Ar), 6.18 (t, 1H, *J* = 8.0Hz, H-Ar), 4.79 (d, 1H, *J* = 21.2Hz, P-CH). ^{13}C NMR (100.6 MHz, CD_3OD): δ 153.26, 147.42, 147.39, 144.81, 144.69, 135.93, 129.74, 129.70, 129.25, 127.34, 126.56, 125.27, 123.91, 122.20, 117.49, 106.22, 61.68 (d, *J* = 130.1 Hz, P-C). ^{31}P-NMR (161.93 MHz, CD_3OD): δ 14.04. HRMS: [$C_{17}H_{15}N_2O_5P$]: m/z = 358.0640, Found $[M-H]^-$ = 357.0635.



57. ((Anthracen-1-ylamino) (4-(trifluoromethyl)phenyl)methyl)phosphonic acid

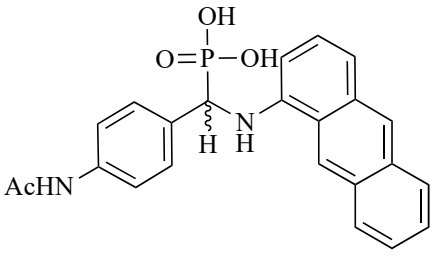
TMSBr (0.128 mL, 0.95 mmol, 5.0 equiv.) was added dropwise to a solution of the aminophosphonate **S57** (90 mg, 0.19 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated in vacuo. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (75 % MeOH/H_2O) to give the phosphonic acid **compound 57** as a yellow solid (55.3mg, 56.4 %). Spectroscopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 9.02 (s, 1H, H-Ar), 8.23 (s, 1H, H-Ar), 8.15 (t, 1H, *J* = 8.0 Hz, H-Ar), 7.91(t, 1H, *J* = 5.6 Hz, H-Ar), 7.81 (d, 2H, *J* = 8.1Hz, H-Ar), 7.51-7.39 (m, 4H, H-Ar), 7.20 (d, 2H, *J* = 8.8Hz, H-Ar), 7.08 (t, 1H, *J* = 16 Hz, H-Ar), 6.143 (d, 2H, *J* = 7.2Hz, H-Ar), 4.82 (d, 1H, *J* = 21.2Hz, P-CH). ^{13}C NMR (100.6 MHz, CD_3OD): δ 148.74, 144.61, 134.23, 132.87, 132.20, 129.64, 129.34, 129.30, 128.60, 128.32, 127.52, 127.23, 126.79, 126.26, 125.58, 125.35, 124.83, 122.65, 120.89, 117.19, 103.79,

61.35 (d, *J* = 129.6 Hz, P-C), ^{31}P-NMR (161.93 MHz, CD_3OD): δ 14.70. HRMS: [$C_{22}H_{17}F_3NO_3P$]: m/z = 431.0862, Found $[M-H]^-$ = 430.0841.



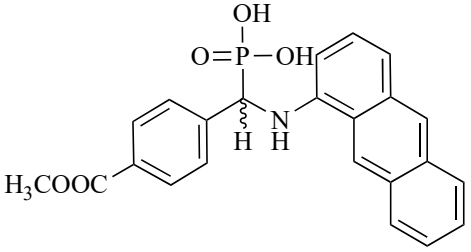
58. ((Anthracen-1-ylamino) (3,5-bis (trifluoromethyl) phenyl) methyl) phosphonic acid

TMSBr (0.075 mL, 0.57 mmol, 3.0 equiv.) was added dropwise to a solution of the aminophosphonate **S58** (100 mg, 0.19 mmol, 1.0 equiv.) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated in vacuo. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (80 % MeOH/H_2O) to give the phosphonic acid **compound 58** as a yellow solid (62.4 mg, 63.1 %). Spectroscopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 9.02 (s, 1H, H-Ar), 8.28-8.14 (m, 4H, H-Ar), 7.96(t, 1H, H-Ar, *J* = 9.2Hz), 7.64 (s, 1H, H-Ar), 7.44-7.42 (m, 2H, H-Ar), 7.25 (d, 2H, *J* = 8.8Hz, H-Ar), 7.06 (t, 1H, *J* = 8.0Hz, H-Ar), 6.07 (d, 1H, *J* = 8.1 Hz, H-Ar), 4.83 (d, 1H, *J* = 20.8Hz, P-CH). ^{13}C NMR (100.6 MHz, CD_3OD): δ 147.97, 144.40, 143.69, 137.54, 134.29, 133.05, 132.38, 129.82, 129.36, 128.65, 127.21, 126.85, 126.39, 125.65, 125.37, 122.65, 121.97, 121.08, 121.19, 117.64, 116.36, 103.59, 61.01 (d, *J* = 129.6 Hz, P-C). 31P-NMR (161.93 MHz, CD_3OD): δ 17.76. HRMS: [$C_{23}H_{16}F_6NO_3P$]: m/z = 499.0802, Found $[M-H]^-$ = 498.0761.



59. ((4-acetamidophenyl) (anthracen-1-ylamino) methyl) phosphonic acid

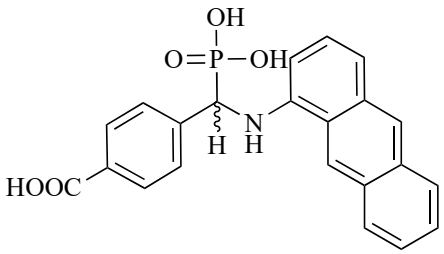
TMSBr (0.115 mL, 0.88 mmol, 4 equiv.) was added dropwise to a solution of the aminophosphonate **S59** (100 mg, 0.22 mmol) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (35 % MeOH/H_2O) to give the phosphonic acid **compound 59** as a yellow solid (40 mg, 41%). Spectrocopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 9.00 (s, 1H, Ar), 8.22 (s, 1H, Ar), 8.16 – 8.08 (m, 1H, Ar), 7.93 – 7.85 (m, 1H, Ar), 7.63 – 7.54 (m, 2H, Ar), 7.45 – 7.31 (m, 4H, Ar), 7.16 (d, *J* = 8.4 Hz, 1H, Ar), 7.05 (t, *J* = 7.9 Hz, 1H, Ar), 6.20 (d, *J* = 7.4 Hz, 1H, Ar), 4.69 (d, *J* = 20.9 Hz, 1H), 2.06 (s, 3H, Me). ^{13}C NMR (100.6 MHz, CD_3OD): 171.52, 144.94, 140.01, 137.19, 134.33, 132.89, 132.23, 129.72, 129.29, 128.67, 127.44, 126.77, 126.28, 125.52, 121.17, 121.00, 116.88, 103.96, 60.36 (d, *J* = 129.8 Hz), 23.75. ^{31}P-NMR (CD_3OD, 161.93 MHz): δ 15.27. HRMS (ESI): Calculated for $C_{23}H_{21}N_2O_4P$ $[M-H]^-$: 419.1161. Found: 419.1184.



60. ((anthracen-1-ylamino) (4-(methoxycarbonyl)phenyl) methyl) phosphonic acid

TMSBr (0.19 mL, 1.45 mmol, 6.5 equiv.) was added dropwise to a solution of the aminophosphonate **S7** (100 mg, 0.22 mmol) in anhydrous DCM/DMF (3 mL, 2:1) at

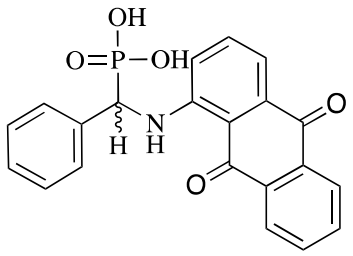
0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL). Methyl group in methyl ester is removed by adding 0.2ml 5M NaOH at 0°C and allowed to warm up to room temperature after 10 minutes. The reaction stops after 5 hours then purified by C-18 reverse phase chromatography (10% MeOH/H_2O) to give the phosphonic acid **compound 8** as a yellow solid (68 mg, 76%). Spectroscopic data: ^{1}H NMR (400 MHz, D_2O) δ 8.99 (s, 1H, Ar), 8.18 (t, *J* = 5.1 Hz, 2H, Ar), 7.88 – 7.75 (m, 3H, Ar), 7.62 (d, *J* = 7.9 Hz, 2H, Ar), 7.51 – 7.41 (m, 2H, Ar), 7.16 (d, *J* = 8.7 Hz, 1H, Ar), 7.12 – 7.03 (m, 1H, Ar), 6.32 (d, *J* = 7.3 Hz, 1H, Ar), 4.75 (d, *J* = 6.9 Hz, 1H, P-CH). ^{13}C NMR (101 MHz, D_2O) δ 175.84, 145.22, 142.97, 134.08, 132.26, 130.69, 128.64, 128.54, 127.54, 127.48, 127.44, 126.46, 126.17, 125.99, 125.42, 119.66, 116.57, 102.80, 58.80 (d, *J* = 128.4 Hz). ^{31}P NMR (162 MHz, Deuterium Oxide) δ 14.93. HRMS (ESI): Calculated for $C_{23}H_{21}N_2O_4P$ [M-H]$^-$: 406.0844. Found: 406.0855.



61. 4-((anthracen-1-ylamino)(phosphono)methyl)benzoic acid

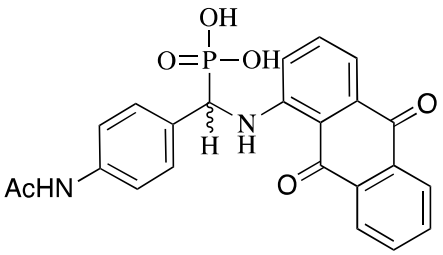
TMSBr (0.19 mL, 1.45 mmol, 6.5 equiv.) was added dropwise to a solution of the aminophosphonate **S7** (100 mg, 0.22 mmol) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL). Methyl group in methyl ester is removed by adding 0.2ml 5M NaOH at 0°C and allowed to warm up to room temperature after 10 minutes. The reaction stops after 5 hours then purified by C-18 reverse phase chromatography (10% MeOH/H_2O) to give the phosphonic acid **compound 8** as a

yellow solid (68 mg, 76%). Spectroscopic data: ^{1}H NMR (400 MHz, D_2O) δ 8.99 (s, 1H, Ar), 8.18 (t, *J* = 5.1 Hz, 2H, Ar), 7.88 – 7.75 (m, 3H, Ar), 7.62 (d, *J* = 7.9 Hz, 2H, Ar), 7.51 – 7.41 (m, 2H, Ar), 7.16 (d, *J* = 8.7 Hz, 1H, Ar), 7.12 – 7.03 (m, 1H, Ar), 6.32 (d, *J* = 7.3 Hz, 1H, Ar), 4.75 (d, *J* = 6.9 Hz, 1H, P-CH). ^{13}C NMR (101 MHz, D_2O) δ 175.84, 145.22, 142.97, 134.08, 132.26, 130.69, 128.64, 128.54, 127.54, 127.48, 127.44, 126.46, 126.17, 125.99, 125.42, 119.66, 116.57, 102.80, 58.80 (d, *J* = 128.4 Hz). ^{31}P NMR (162 MHz, Deuterium Oxide) δ 14.93. HRMS (ESI): Calculated for $C_{23}H_{21}N_2O_4P$ $[M-H]^-$: 406.0844. Found: 406.0855.



62. (((9,10-dioxo-9,10-dihydroanthracen-1-yl) amino) (phenyl)methyl) phosphonic acid

MSBr (0.155 mL, 1.19 mmol, 5 equiv.) was added dropwise to a solution of the aminophosphonate **S62** (100 mg, 0.237 mmol) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (20% MeOH/H_2O) to give the phosphonic acid **compound 62** as a purple solid (89 mg, 95%). Spectroscopic data: ^{1}H NMR (400 MHz, D_2O) δ 7.87 (s, 1H, Ar), 7.57 (s, 2H, Ar), 7.43 (s, 3H, Ar), 7.37 (s, 1H, Ar), 7.16 (d, *J* = 7.7 Hz, 1H, Ar), 6.98 (d, *J* = 7.6 Hz, 1H, Ar), 6.86 (d, *J* = 7.3 Hz, 2H, Ar), 4.68 (s, 1H, P-CH). ^{13}C NMR (101 MHz, D_2O) δ 185.01, 184.87, 140.92, 135.29, 134.67, 133.29, 131.68, 128.29, 128.02, 126.73, 126.21, 120.33, 116.27, 112.52, 57.71 (d, *J* = 131.9 Hz). ^{31}P NMR (162 MHz, D_2O) δ 14.21. HRMS (ESI): Calculated for $C_{21}H_{16}NO_5P$ $[M-H]^-$: 329.0766. Found: 392.0762.



63.((4-acetamidophenyl) ((9,10-dioxo-9,10-dihydroanthracen-1yl) amino) methyl) phosphonic acid

TMSBr (0.115 mL, 0.88 mmol, 4 equiv.) was added dropwise to a solution of the aminophosphonate **S63** (100 mg, 0.22 mmol) in anhydrous DCM/DMF (3 mL, 2:1) at 0°C and the mixture held at this temperature for 30 minutes. The solution was then allowed to warm to room temperature and concentrated *in vacuo*. The resultant residue was then diluted with water (2.5 mL), neutralized (aq. 1 M NaOH) and purified by C-18 reverse phase chromatography (35 % MeOH/H_2O) to give the phosphonic acid **compound 63** as a yellow solid (40 mg, 41%). Spectroscopic data: ^{1}H-NMR (400 MHz, CD_3OD): δ 9.00 (s, 1H, Ar), 8.22 (s, 1H, Ar), 8.16 – 8.08 (m, 1H, Ar), 7.93 – 7.85 (m, 1H, Ar), 7.63 – 7.54 (m, 2H, Ar), 7.45 – 7.31 (m, 4H, Ar), 7.16 (d, J = 8.4 Hz, 1H, Ar), 7.05 (t, J = 7.9 Hz, 1H, Ar), 6.20 (d, J = 7.4 Hz, 1H, Ar), 4.69 (d, J = 20.9 Hz, 1H), 2.06 (s, 3H, Me). ^{13}C NMR (100.6 MHz, CD_3OD): 171.52, 144.94, 140.01, 137.19, 134.33, 132.89, 132.23, 129.72, 129.29, 128.67, 127.44, 126.77, 126.28, 125.52, 121.17, 121.00, 116.88, 103.96, 60.36 (d, J = 129.8 Hz), 23.75. ^{31}P-NMR (CD_3OD, 161.93 MHz): δ 15.27. HRMS (ESI): Calculated for $C_{23}H_{21}N_2O_4P$ $[M-H]^-$: 419.1161. Found: 419.1184.

7.2.4. Downstream protocols of chemical synthesis

7.2.4.1. Thin layer chromatography

Thin layer chromatography (TLC) was used in different steps of organic synthesis to separate chemical mixtures and rapidly determine the components in a reaction mix or

fractions eluted from columns. TLC can be used to monitor the progress of reaction, identify the compound percentage from a mixture, or determine the purity of substances. TLC is performed on a silica coated foil plate (Sigma), with the sample spotted onto the plate using a capillary. After application of the sample, the TLC plate was placed in a mobile phase (solvent or solvent mixture), Different analytes ascend the plate at different rates to achieve the separation (Vogel and Furniss, 1989).

The separation of components can be observed by UV light (wavelength 254 nm). The TLC plate was then stained by Potassium Permanganate solution (1.5g $KMnO_4$, 10g K_2CO_3, and 1.25mL 10% NaOH, dissolved in 200mL water) or p-anisaldehyde stain (135 mL of absolute ethanol, 5 mL of concentrated sulfuric acid, 1.5 mL of glacial acetic acid, 3.7 mL of p-anisaldehyde) so that after drying, the spots of analytes would be visible.

7.2.4.2. Purification of small molecule compounds

All synthesized small molecule compounds were purified using column chromatography, which is a fundamental method to purify individual compounds from a mixture. For compounds which have high solubility in organic solvent, the columns use silica gel as the stationary phase, the ratio between the silica gel and the analyte is around 30:1. The solvent system of mobile phase was determined by TLC beforehand. Silica gel was mixed with the starting solvent or solvent mixture then placed into the column with cotton wool and silicon dioxide on the bottom. For liquid analyte mixtures, the sample was loaded directly on the top of packed silica gel, the analyte mixture in solid phase would be dissolved by the minimum volume of the starting solvent or solvent mix before loading onto the packed column. For analytes that were insoluble in starting solvent, the sample was dissolved completely in appropriate solvent(s) and mixed with silica gel. The sample-silica gel mixture was then dried gently using a rotary evaporator and loaded on the column carefully. After loading, elution from the column is started by adding pressure, fractions were separated and analysed by TLC and mass

spectrometry (MS) (Burke, Germany), and pure fractions were collected in a round bottom flask and dried through a rotary evaporator then a high vacuum pump.

Compounds with high polarity or high aqueous solubility were purified by Reversed-phase chromatography using SNAP C18 prepacked cartridges (Biotage, Sweden), an octadecyl carbon chain (C18) bond silica based column. Typically, the mobile phase of a reversed phase column used a mixture of water and methanol. Firstly, the column was washed with 400ml of methanol and prepared by running 100ml of distill water through it. The sample was dissolved in the minimum volume of water (1-2ml) and loaded onto the column by syringe. The mobile phase had a gradient of methanol concentration. Fractions were analysed by TLC and MS, pure fractions were collected and dried using a dry ice bath rotary evaporator, and water was removed by freeze drying (Biobase, Germany).

7.3. Biology and biochemistry

7.3.1. Bioinformatics

7.3.1.1. Molecular docking

Molecular docking is one of the most widely used virtual screening techniques for drug discovery. By computationally docking the small molecule candidate to a potential binding site of the macromolecule acceptor, the potential binding position and binding energy can be predicted and evaluated. Although many limitations still exist, for example, rigid docking does not normally consider the flexibility of a protein structure and nor the solvent system, docking is still a useful tool for drug development. In this study, we perform molecular docking of α-aminophosphonate compounds into the catalytic pocket of TcTS, predict the binding energy and bound structures, which can provide a reference point for experimental drug candidate screening.

A protein model of TcTS was first prepared by AutoDockTools (Morris *et al*., 2009) using an unliganded TcTS model (PDB: 1MS3), all hydrogens were added to the model and the docking grid box was limited to the catalytic site, and Tyr119 and Trp312 were set to be flexible residues. Three-dimensional models of ligands were built using Phenix eLbow (Moriarty *et al*., 2009) or JLigand (Lebedev *et al*., 2012), based on the SMILE string and/or on a 2D model built by PerkinElmer Chemdraw.

Molecular docking experiments can be run using AutoDock Vina (Trott and Olson, 2010) after the preparation of both the ligand and protein models, AutoDock Vina will screen the possible binding positions and calculate the predicted binding energy as the docking score. At the end of the process, the top 10 high ranked positions will be listed and the binding model presented by PyMol (Janson *et al*., 2016).

7.3.1.2. Multiple sequence alignment

Bioinformatics methods were applied to examine the allosteric binding site by comparing different strains of TcTS, and to evaluate the uniqueness of the allosteric binding site among different sialidases in order to evaluate the specificity of aminophosphonate. Initially, TcTS sequence from different strains were obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Sequence alignment in each strain and among different strains were performed by Expresso, a protein sequence alignment server using structural information (Armougom *et al*., 2006). The conservation of residues in the catalytic and allosteric sites was presented by ESPript (http://espript.ibcp.fr).

7.3.2. Cloning and transformation

7.3.2.1. Transformation

Vectors containing inserts which encode recombinant proteins were transformed to appropriate cell lines for amplitude or expression. Plasmids were first transformed into the cloning cell line NEB5α. The transformation experiments followed a typical heat shock transformation protocol (van der Rest *et al*., 1999), All plasmids were dissolved in distilled water, and diluted to around 125 ng/μl (DNA concentration was detected by spectrometer). Chemical competent cells (10 μl) were thawed on ice for 10min. 1-5μl of plasmid was added into the competent cell and mixed gently before incubation on ice for 30mins. The pre-incubated transformation mixture of NEB5α was placed in a 42℃ water bath for exactly 30 seconds and transferred onto ice immediately. After transformation, 900 μl of SOC media was added into the reaction mixture and recover at 37℃ for 1 hr with 180-200 rpm shaking. Then 100-500 μl recovered culture was spread on a LB agar plate (containing appropriate antibiotics) and incubate at 37℃ overnight to obtain the single colonies.

1-5 single colonies of NEB5α were picked and grown overnight in 15ml LB media contacting 1mM ampicillin. Cells were harvested by centrifugation (8500g, 10min), and the amplified plasmid extracted using a GeneJET plasmid miniprep kit (Thermal scientific, UK). 10μl of extracted plasmid was sent for sequencing by SourceBioscience, UK.

After the sequences were confirmed, the plasmids were transformed into the expression cell line BL21 (DE3), parameters were chosen following the advice of the manufacturers of the different competent cells.

7.3.2.2. Site directed mutagenesis

Site directed mutagenesis was performed following a general protocol (Weiner *et al*., 1994). The E. *coli* strain containing the gene sequence of the recombinant protein was first resuscitated overnight in 15ml LB media with appropriate antibiotics. The plasmid DNA was extracted using a GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). DNA primers with the mutated site mismatch were designed using the QuickChange primer design program (access from http://www.genomics.agilent.com /primerDesignProgram.jsp) and synthesized by Invitrogen, UK. The Polymerase chain reaction (PCR) mixture was prepared in a thin wall PCR tube using the recipe shown in Table 2.2 and mixed gently. The PCR mixture was then placed in the PCR machine (MJ Research), and the reaction was run as shown in table 2.3.

Table 7.2. PCR mix of mutagenesis using *Pfu* polymerase

10X *Pfu* Buffer with $MgSO_4$	5 µl (0.2 mM of each)
dNTP Mix, 10 mM	1 µl (0.2 mM of each)
Forward primer	1.0 µM
Reverse primer	1.0 µM
Template DNA	1µg
Pfu DNA Polymerase	1.25-2.5 U
Water	Top up to 50µl
Total	50µl

Table 7.3. Cycling parameters for the site-directed mutagenesis PCR reaction

Segment	Cycles	Temperature	Time
1	1	95°C	30 seconds
2	12–18	95°C	30 seconds
		55°C	1 minute
		68°C	1 minute/kb of plasmid length*

PCR products were examined by DNA gel electrophoresis. In next step, Methylated and hemimethylated DNA was removed by DpnI digestion. A digestion reaction mix was prepared on ice, the reaction mix then transferred into a 37°C water bath for 30min and chilled on ice. Transformation was performed by the heat-shock method immediately after the digest reaction was finished. The result of mutagenesis was examined by DNA sequencing (Source Bioscience, UK).

Based on the crystal structure, a novel allosteric site was found through aminophosphate inhibitor binding. To interpret the authenticity of our observation from the soaked crystal structure, site direct mutagenesis of residues around the allosteric site was performed. Three mutations, **R249A, R275A, R577A** were performed through a standard site-direct mutagenesis protocol (Liu and Naismith, 2008). Alanine was chosen to remove the side chain of these three Arginines while maintaining the stability of the C-α chains. Another mutagenesis, **D59A** was generated to interpret lactose binding with and without inhibitors.

The method relies on two complement PCR primers, which incorporate both the selection site and the desired single base substitutions at the mutant site. Forward and Reverse primers of all four mutations (Appendix 8.5) were designed with a single mismatch for the mutated residue.

Table 7.4. Primers for TcTS mutagenesis

Name	Sequence
TcTS_R249A_F	accagacggcgggcatagtcaactcgagtgtttatga
TcTS_R249A_R	tcataaacactcgagttgactatgcccgccgtctggt
TcTS_R272A_F	ccccacacagctgagagcgtgccgacag
TcTS_R272A_R	ctgtcggcacgctctcagctgtgtgggg
TcTS_R577A_F	gttgtgccagacgaggcgacgcctgacatctc
TcTS_R577A_R	gagatgtcaggcgtcgcctcgtctggcacaac
Name	**Sequence**
TcTS_D59N_F	tgagggagttgttaaaggatgtttcgtagcgagcgt
TcTS_D59N_R	acgctcgctacgaaacatcctttaacaactccctca

Table 7.5. Reaction mix of dpnI digestion

Components	Volume
Water	15 µl
10X FastDigest or 10X FastDigest Green Buffer	2 µl
PCR product	2 µl (up to 1 µg)
FastDigest dpnI	1 µl
Total	20µl

7.3.3. Expression and purification of recombinant protein

7.3.3.1. Expression of recombinant protein by IPTG induction

This section describes the expression of recombinant protein via IPTG induction, for either TcTS or SpNA. After an expression vector was transformed into an expression *E.*

coli strain and spread on agar plates with appropriate antibiotics (1 μM ampicillin for TcTS and 1 μM kanamycin for SpNA), a single colony was picked and grown at 37°C overnight in 15 ml of LB medium with addition of appropriate antibiotic in a temperature-controlled shaking incubator. The overnight primary culture was used to incubate a larger scale secondary LB culture (250-1000mL) to which had been added the appropriate antibiotics (1 μM ampicillin for TcTS and 1 μM kanamycin for SpNA). Cells were grown at 37°C with shaking at between 180 to 200 rpm. The induction of over-expression of the recombinant protein was achieved by adding isopropyl thiogalactoside (IPTG) during the exponential phase of bacterial growth (OD_{600} = 0.6-0.8) and incubate at 25°C for 12-16 hr with 180 to 200 rpm shaking.

7.3.3.2. Auto induction

Instead of IPTG induction, recombinant protein from vectors that contain genes that relate to sugar metabolism, for example T7 and pTrc promoter, underwent auto-induction. The primary culture was the same as IPTG induction. The secondary culture was started by adding 5ml of primary culture into 1L TB media with appropriate antibiotics and incubated at 37°C with 200rpm shaking for 2 hours, then the temperature was lowered to 16°C and incubation continued for 48-60 hours.

7.3.3.3. Harvest and lysis of cells

Induced cells were harvested by centrifugation at 4°C, 8,000 g for 30 minutes and stored at -80°C. To lyse the cells, a cell pellet was thawed and re-suspended in His A buffer (table 7.6). The cell suspension then lysed using a cell disruptor (Branson, USA) for 6 cycles (10sec per cycle with 10min gap between cycles). The solution was centrifuged to remove insoluble cell debris at 4°C, 70,000 g for 30mins.

7.3.3.4. Protein purification

All recombinant proteins in this study are His-tagged, thus the most common method for protein purification is Ni-affinity chromatography and further purification was achieved through ion exchange chromatography. The columns used for each method and the buffers are listed in table 7.6. All purifications were performed using a ÄKTA purifier protein purification system (GE Healthcare)

Table 7.6. Methods, columns and buffers used in protein purification

Protein	$TcTS/TcTS_{mut}$	SpNA
Method	Nickel affinity chromatography	
Column	His-trap FF 1ml/5ml (GE Healthcare)	
Buffer	Binding buffer (His A): 50mM Tris, 500mM NaCl, 20mM Imidazole, pH 7.4	
	Elution buffer (His B): 50mM Tris, 500mM NaCl, 500mM Imidazole, pH 7.4	
Method	Anion-exchange chromatography	
Column	Hi-trap Q FF 1ml/5ml (GE Healthcare)	
Buffer	Binding buffer (Anion A): 50mM Tris, pH 9	
	Elution buffer (Anion B): 50mM Tris, 500mM NaCl, pH 9	
Method	Gel filtration	
Column	Superdex 200 10/300 GL (GE Healthcare)	
Buffer	20mM Tris, pH 7.5	N/A

7.3.3.5. Buffer exchange

Four buffer-exchanging methods were used in this study. Amicon-Ultra (Merck Millipore) centrifugal filters were used to enrich purified protein, several cycles of spinning dilutions-enrichment were repeated utilizing the replacement buffer, until an equivalent of 1000× dilution (4 cycles of 25× dilution) of the original buffer was attained. Alternatively, protein sample were buffer exchanged by PD-10 gravitational desalting column (GE-Healthcare). Briefly, a PD-10 gravitational column was pre-equilibrated with 25ml of replacement buffer. Up to 2.5 ml of protein sample was loaded on the top of the column. When the protein sample had been fully absorbed by the column, 3.5 ml of replacement buffer was added then 3.5ml of protein solution was eluted in the desired buffer. For small volumes of protein, protein desalting spin columns (10,000Da cut off, Thermo Scientific) were used. The spin column was first equilibrated with replacement buffer through 3 wash cycles (400 µl per cycle), and the excess liquid was removed by 1 min centrifugation at 1500 *g*. The protein sample was then placed on the center of the resin bed and the column was placed above a new collection tube, after which the column was spun at 1500 *g* for 2 minutes to collect the buffer-exchanged sample.

To exchange the buffer more completely, classical dialysis was performed by using Slide-A-Lyzer™ G2 Dialysis Cassettes (10,000Da cut off, Thermo Scientific). Dialysis cassettes were first prepared by soaking in the replacement buffer or ddH_2O for 2 min. Samples were added into the cassette by pipetting or using a syringe with a needle, followed by removal of excess air in the cassette. The cassette was placed in a replacement buffer of 200× higher volume than protein sample. The dialysis process was performed at 4°C in a cold room with gentle stirring, the dialysis buffer was changed 3-4 times in 24 hours to make an 8,000,000 times dilution of the original buffer.

7.3.3.6. Concentrator and concentration measurement

While protein samples were collected from purification or buffer exchange, the concentration was increased using Amicon-Ultra centrifugal filters. Firstly, centrifugal filter was equilibrated by adding the desired buffer and spinning for 5 min. Protein samples were then placed on the center of the filters and spun at 3500 *g* until the appropriate volume was reached. Samples were finally collected by pipetting carefully. The concentration of protein samples was determined by measuring the absorbance at 280nm using a BioSpectrometer (Eppendorf).

7.3.4. Downstream protocols of nucleotide and protein analysis

7.3.4.1. DNA gel electrophoresis

DNA samples include plasmids and PCR products were detected by agarose gel electrophoresis. Agarose gels were made by dissolving 1% (w/v) agarose in 1× TAE buffer (40mM Tris-acetate, 1mM EDTA, pH8.0) and addition of 1 in 20,000 dilution of 10mg/ml ethidium bromide solution for visualization of the DNA. 1µl of 6× DNA loading dye was mixed with 5 µl DNA sample before loading onto the agarose gel. Electrophoresis was performed using horizontal flatbed chambers (BioRad) at a constant current of 80 mA. For DNA size estimation, the 1Kb Plus DNA ladder (Figure 7.1) was utilized. Gels were imaged using a Chemidoc-it^2 UV transilluminator (UVP).

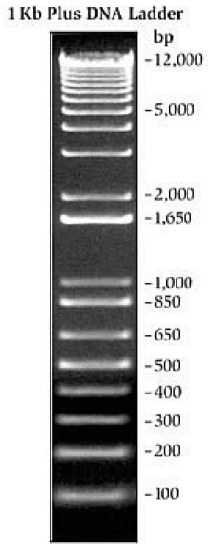


Figure 7.1. Markers used for DNA size estimation (1Kb plus DNA ladder, BioRad)

7.3.4.2. DNA purification

PCR products were purified using a MiniElute PCR purification kit (Qiagen) to remove protein contaminants in the sample. In cases where PCR reactions produced multiple bands on a gel, the whole PCR reaction mixture was electrophoresed, and the desired band was excised for DNA extraction using MiniElute gel extraction kit (Qiagen). In both purification and gel extraction, DNA samples were finally eluted with 10 µl sterile water.

7.3.4.3. SDS-Page electrophoresis

In this study, protein samples were estimated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) using a Novex Mini gel system (Invitrogen) using 4% polyacrylamide stacking gel (Appendix 8.3), and 10%, 12% and 15% polyacrylamide resolving gel (Appendix 8.3). Protein analytes were mixed with 2× SDS-PAGE loading buffer (Appendix 8.3) and heated at 100°C for 10 min. Generally, 5-15 µl samples (with less than approximately 25µg protein) were loaded into each lane. Electrophoresis were performed in 1× Tris-Glycine running buffer

(Dilute from 10× stock, Appendix 8.3), using the constant current setting of 60-80 mA until the bromophenol blue dye ran off the gel. The Gel was extracted from the cassette then stained with Coomassie Blue (Appendix 8.3) for 30-60 min after being heated in microwave for 40 sec. The excess stain was then removed using de-stain solution (Appendix 8.3) overnight and the gel was imaged using a Chemidoc-it^2 UV transilluminator. PageRuler Unstained (Figure 7.2A) and Prestained (Figure 7.2B) protein ladder (Thermo Scientific) were used estimate the size of proteins in the sample.

.

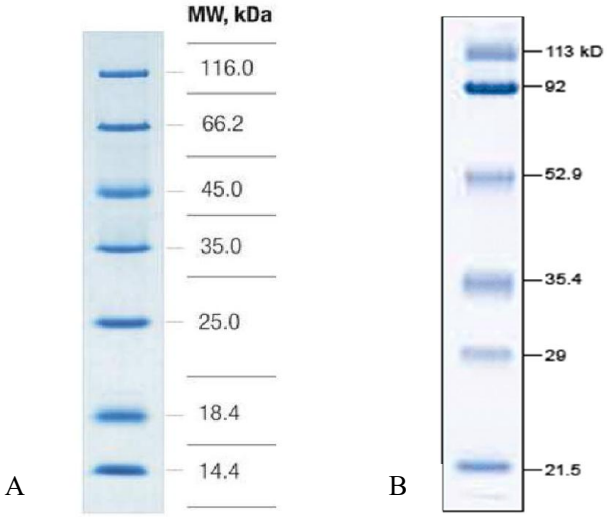


Figure 7.2. Markers used for protein size estimation: A: Unstained protein ladder (Thermo Scientific), B: Prestained protein ladder (Thermo Scientific)

7.3.5. TcTS/Sialidase activity and inhibition assay

The sialidase activity of TcTS and sialidases was determined using 2′-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid (MuNANA) as substrate. MuNANA is a non-natural substrate which has been widely used in determination of enzymic activity of sialic acid hydrolases or transferases (Potier *et al*., 1979). When the Oxygen-linkage between sialic acid 4-Methylumbellifrerone (4-Mu) is cleaved by sialidase or trans-sialidase, the shed 4-Mu can generate a strong fluorescence signal, which is significantly higher than the self-fluorescence which is generated by MuNANA molecules (Marathe *et al*., 2013) (Figure 7.3.).

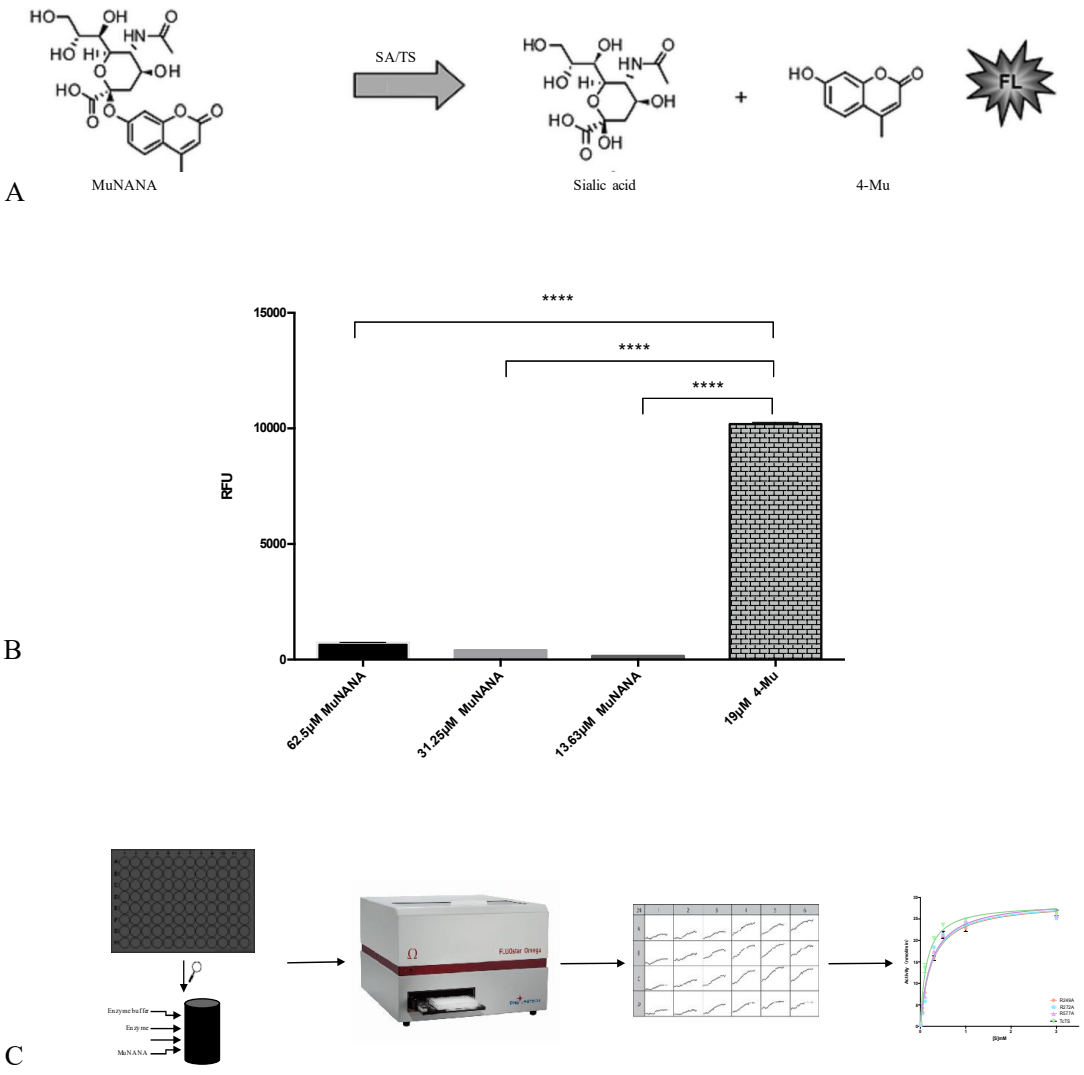


Figure 7.3. A: Principle of fluorescence SA enzyme assay using MuNANA (Grienke *et al*., 2012) B: Fluorescence generation by different concentrations of MuNANA, compared with 19 µM 4-Mu sodium salt in enzyme buffer in 100 µl reaction (Marathe, *et al.* 2013) C: Process of fluorescence SA assay.

In this study, we performed SA activity and inhibition assays in Cellstar 96 well solid black plates (Greiner). The reaction system, with total volume of 100 µl, was prepared by addition of assay buffer (Appendix 8.5), enzyme stock solution and inhibitor (only in case of inhibition assays). The reaction mixture was pre-incubated at appropriate

temperature for at least 10 min. The stock solution of MuNANA, which has been pre-incubated at the assay temperature, then was added into the reaction mixture carefully. The plate then quickly placed into a FlOUstar Omega multimode microplate reader (BMG labtech) and reading started with the excitation wavelength at 360 nm and emission wavelength at 460nm. The assay was normally read for approximately 10min and fluorescence signal generated by hydrolyzed 4-Mu was collected and plotted. All assays were repeated three times independently. The raw data were analyzed by MARs data analysis software (BMG labtech), a linear region of the raw data curves was picked and the slope of the curve was calculated using unit RFU/min, which represents the activity of the enzyme.

Further analysis of the raw data was performed by GraphPad Prism software.

K_m/V_{max}/kcat

To characterize the enzyme, the K_m and Vmax are the first parameters to be determined. K_m/V_{max}/k_{cat} assay was run with a constant enzyme concentration and various substrate concentrations. The enzyme activity at these different concentrations of substrate was recorded and a non-linear fit was performed. The K_m and Vax were calculated using following equations, in which RFU stands for the fluoresce unit, and [Er] means the concentration of enzyme.

$$Activity\ (RFU/min) = \frac{V_{max} \times x}{K_m + x}$$

$$kcat = V_{max}[Er]$$

IC_{50}

For high throughput screening of the inhibitor candidates and analysis of selectivity of inhibitors, half maximal inhibitory concentration (IC_{50}) were tested. The IC_{50} assay of TcTS and sialidases records enzyme activity with a constant concentration of enzyme and substrate, and various concentration of inhibitor. Enzyme activity without inhibitor is regarded as 100% and the percentage of enzyme activity remaining after the addition

of inhibitor can be calculated. IC_{50} is calculated using the following equation:

$$\%\ activity = min + \frac{max - min}{1 + 10^{(logIC_{50} - \log[I])hillslope}}$$

where *min* and *max* are the minimum and maximum limitations of the non-linear curve fitting, Hill slope, also known as hill coefficients, is the slope of the non-linear fitting, and it can be interpreted to determine the inhibition mechanism and binding stoichiometry (Prinz, 2010). The time-dependent inhibition assay also follow the procedure but with various pre-incubation times of the enzyme-inhibitor mixture.

Mode of inhibition and K_i

The assay which detects the mode of inhibition is similar to the K_m/Vmax assay, but the K_m and Vmax are detected in the presence of different concentrations of inhibitor. The results were fitted by the following models:

Competitive inhibition:

$$K_{m(Obs)} = K_m \frac{1 + [I]}{Ki}$$

Non-competitive inhibition:

$$V_{max}inh = \frac{Vmax}{1 + I/Ki}$$

Mixed inhibition:

$$V_{\max(App)} = \frac{V_{max}}{1 + I/(\alpha \times Ki)}$$

$$K_{m(App)} = Km * (1 + I/Ki)/(1 + I/(\alpha * Ki))$$

Un-competitive inhibition:

$$VmaxApp = Vmax/(1 + I/\alpha Ki)$$

$$KmApp = Km/(1 + I/\alpha Ki)$$

The model fitting results were compared and the one with the least standard error was selected as the optimal inhibition model.

7.3.6. Thermodynamic study

The thermodynamic characterization of interactions between TcTS and either inhibitor or substrate can be investigated using isothermal titration calorimetry (ITC). Uniquely, ITC can detect both binding constants and thermodynamic parameters of the interaction between two or more molecules in one experiment without any labelling of the samples.

ITC experiments in this study were performed using a NanoITC instrument (TA instrument). Both protein (acceptor) and ligand molecules were dissolved in the same buffer by dialysis or another buffer exchange method. Both protein and ligand were pH matched and degassed before the experiment. ITC assay was performed by placing the protein into the sample cell and the ligand in the syringe. Experimental data was analyzed by NanoAnalysis software (TA instrument) and the isothermal parameters ΔH and ΔS, and the binding constant K_d were calculated.

The thermodynamic characteristics of the interaction between TcTS and α-aminophosphonate inhibitor were determined by isothermal titration calorimetry (ITC). A purified TcTS sample was dialyzed against the ITC assay buffer (PBS, 50mM NaCl, pH 7.5), then concentrated to about 0.11mM. The inhibitor sample was dissolved into the same buffer. After degassing, the TcTS and inhibitor samples were loaded into the sample cell and syringe of the ITC instrument respectively. The ITC experiment was run at 20°C with stirring at 250 rpm. 30 titrations were measured with a gap of 10 mins between two titrations. The result was analyzed by NanoAnalysis software (manufacturer) and isothermal parameters including ΔH and ΔS and the binding constant K_d were calculated.

7.3.7. Protein crystallography

7.3.7.1. Crystallization of recombinant TcTS

The crystallisation of TcTS included robot screening (Phoenix nano dispenser, ARI), which used sitting drop vapor diffusion; performed in MRC 96-well crystallization plate (molecular dimensions, UK) and crystallization screens were sourced from Molecular Dimensions, UK. Manual crystallization and co-crystallization trials were performed using hanging drop vapor diffusion in 24-well XRL Plates with round CrystalClene™ coverslips (Molecular Dimensions, UK). All chemical, without further purification, were directly applied for crystallization trials

TcTS has been crystallized, however, due to the complexity and flexibility of the protein, it is still not easy to obtain good quality single crystals even from known, published, conditions. To ascertain possible new crystallization conditions for TcTS, the commercial screens JCSG screen, PEG screen, and PGR screen (Molecular Dimensions, UK) were set in 96-well plate using Phoenix crystal robot (ARI instrument), 300nl of purified TcTS at 18 mg/ml, was added into each well, with 1:1, 1:2, 2:1 ratio of protein:mother liquor. Manual screening based on the published condition (**Condition A**: 100mM Tris-HCl, 20% PEG4000, 5% isopropanol, pH 7.5) was performed in 24-well plate, with pH range between 7.0 to 8.0, the range of PEG4000 percentage between 10% to 30%, and different buffer condition including Tris-HCl, HEPES and PBS. The present of salt can help the formation of crystal nuclei, therefore a salt screen has also been carried out by adding a wide range of salts based on the published crystal condition (NaCl, $MgSO_4$, $CuSO_4$, $NiCl_2$).

Unfortunately, no hit was found in all the commercial screening and salt screening experiments. Microcrystals start growing in condition A after two weeks, but fail to grow to suitable size for X-ray diffraction after months. A few single crystals were grown in **condition B**: 100mM HEPES, 20% PEG4000, 5% isopropanol, pH7.5. To

obtain larger amount of high quality single crystals, micro and macro seeding has been performed in both condition A and condition B. In micro seeding, microcrystals were transfer into micro centrifuge tube with 100µl of mother liquor, and broken down by vortexing. The seed liquor was diluted by mother liquor in ratios of 1:1, 1:5, 1:10. 1µl of diluted seed solution was then mixed with equal volumes of 18mg/ml purified TcTS. Unlike micro seeding, macro seeding uses microcrystals directly as seeds, the micro crystals were transferred into drops with concentrated protein. Single crystals with appropriate size were grown in condition B after few days to weeks.

7.3.7.2. Crystal soaking and co-crystallization

Good inhibitor candidates (**48, 51, 53, 54, 59, 61, 62** and **63**) from structure-activity screening described in **Chapter 3** (IC_{50}<0.3mM), known inhibitors DANA and Rhein (reported IC_{50}=56µM) (Arioka *et al*., 2010) were used to attempt crystal soaking with TcTS single crystals. For water-soluble compounds, including DANA, lactose, sialic acid and α-aminophosphonate compounds, small molecule compounds were prepared in final solutions at 5-10mM in crystallization mother liquor. Rhein is only partly soluble in aqueous solution, thus, 10% (v/v) of DMSO was added into the crystallization liquor and the final concentration of Rhein was 1 µM. For all the ligands, 1µl of stock solution was added onto a 1µl drop containing a single crystal. The soaked crystals were stored in hanging drops at room temperature before being flash-cooled using liquid nitrogen.

Co-crystallization was performed with compound **59**, purified TcTS (18mg/ml) was pre-incubated with 5mM of compound at room temperature for 30mins. Crystallization trials using the JCSG, PEG and PGR screens were set up by the crystallography robot (Phoenix) at 1:1, 1:2, 2:1 protein: mother-liquor ratio. Hits were observed in condition B12 (0.2 M Potassium citrate tribasic monohydrate, 20 % w/v PEG 3350), G8 (0.15 M DL-Malic acid, 20 % w/v PEG 3350), G11 (2.0 M Ammonium sulfate, 0.1 M BIS-Tris, pH 5.5) in JCSG screen. These initial co-crystallization conditions were optimized

using 24-well plates. A single crystal was found in condition (0.15 M Potassium citrate tribasic monohydrate, 20% w/v PEG 3350).

Good inhibitor candidates (IC_{50} < 0.3 mM) from structure-activity screening described in subsequent chapters, the known inhibitors DANA, lactitol and Rhein (Arioka *et al*., 2010) were used in crystal soaking trials with TcTS single crystals. Substrate, such as SA and lactose, were also applied in soaking experiment in order to interpret the mechanisms of inhibitors. For water-soluble compounds, such as DANA, lactose, sialic acid and α-aminophosphonate compounds, stock solutions of the small molecule compounds were made at 5-10mM in crystallization mother liquor. Rhein is only partly soluble in aqueous solution, thus, 10% (v/v) of DMSO was added into the crystallization liquor and the final concentration of rhein was 1 μM. For all the ligands, 1μl of stock solution was added onto 1μl drop containing single crystal. The soaked crystals were placed as hanging drops at room temperature before being frozen by liquid nitrogen.

Co-crystallization was also performed with compound **59,** selected due to having the best inhibition effect. Purified TcTS (18mg/ml) was pre-incubated with 5mM of compound at room temperature for 30mins. JCSG, PEG, and PGR screenings were performed by the crystallography robot (Phoenix) at 1:1, 1:2, and 2:1 protein-mother liquor ratio. Hit conditions were optimized by hang-drop, vapor diffusion method in 24 well plate.

7.3.7.3. Data collection, processing and refinement

Crystals were cryo-protected with 30% glycerol (w/v) for a few seconds then transfer to liquid nitrogen. All data were collected at 100K on the in-house X-ray diffraction instrument (Rigaku) or at the synchrotron (beam I03, I04, Diamond light source, Oxford). Data processing was performed by either iMosflm (Battye *et al*., 2011),

HKL2000 (Otwinowski *et al*., 1997), or Xia2 (Winter *et al*., 2013). The space group of most crystals was determined as monoclinic (P1 2_1 1), but triclinic (P1) and orthorhombic (P 2_1 2_1 2_1) were also found in some datasets. The data were scaled using SCALA. The Rmerge is calculated by following equation:

$$R_{merge} = \sum_{hkl}\sum_{j=1}^{N}\left|I_{hkl} - I_{hkl}(j)\right| / \sum_{hkl}\sum_{j=1}^{N}\left|I_{hkl}(j)\right|$$

Since the structure of TcTS has been solved, previous published structure 1MS5, 1MR5, 1S0I (Amaya *et al*., 2004; Buschiazzo. *et al*., 2002) were directly used as models for molecular replacement using Balbes (Long *et al*., 2008), CCP4 Phaser (McCoy *et al*., 2007) or Phenix Phaser MR (Bunkoczi *et al*., 2013) using the soaked crystal data. The model built by molecular replacement was then refined by Phenix Refine (Afonine *et al*., 2012) or REFMAC5 (Murshudov *et al*., 2011), the refined model was manually fitted and improved using Coot, as well as ligand and solvent building and fitting (Emsley and Cowtan, 2004). Coordinate and crystallographic information file (.cif files) of inhibitors and other ligands were created by Phenix elbow (Moriarty *et al*., 2009).The R_{factor} for refinement is calculate by following equation:

$$R_{factor} = \sum_{hld}\left|\left|F_{obs}(hkl)\right| - \left|F_{cal}(hkl)\right|\right| / \sum_{hld}\left|F_{obs}(hkl)\right|$$

R_{free} was calculated as defined by Brunger (1992) for 5% of the total data. The refined model was validated by Molprobity (Chen *et al*., 2010).

Chapter 8. Appendix

8.1. Cloning vectors

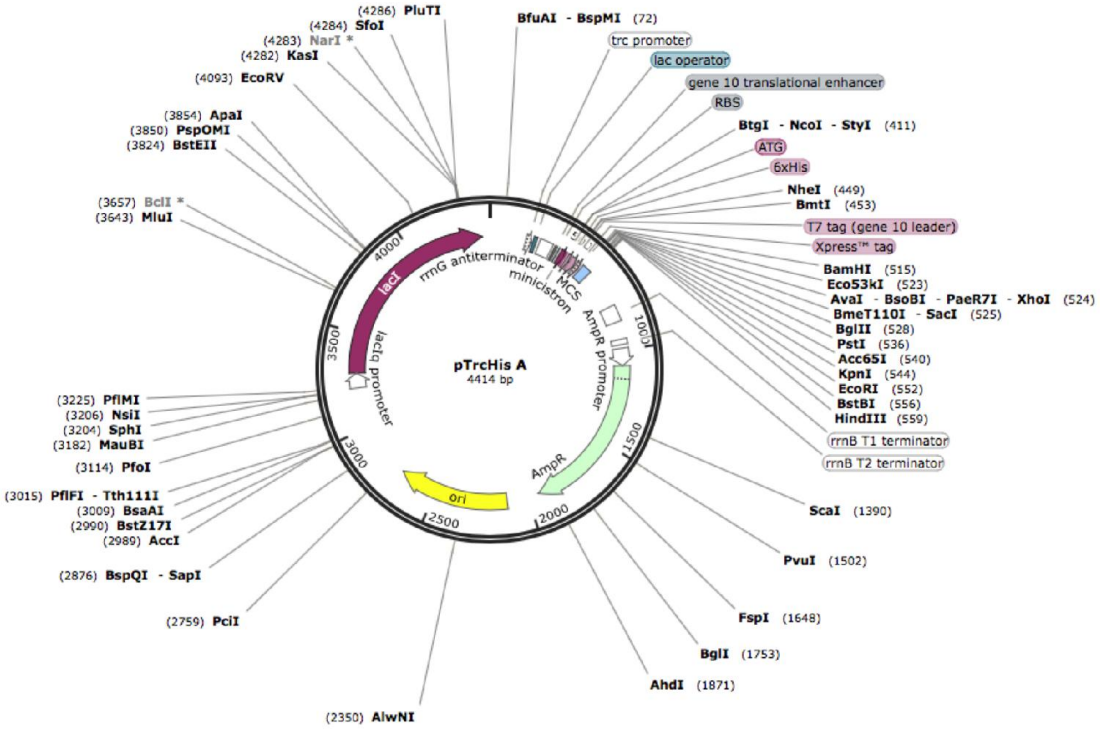


Figure 8.1. pTrcHisA vector

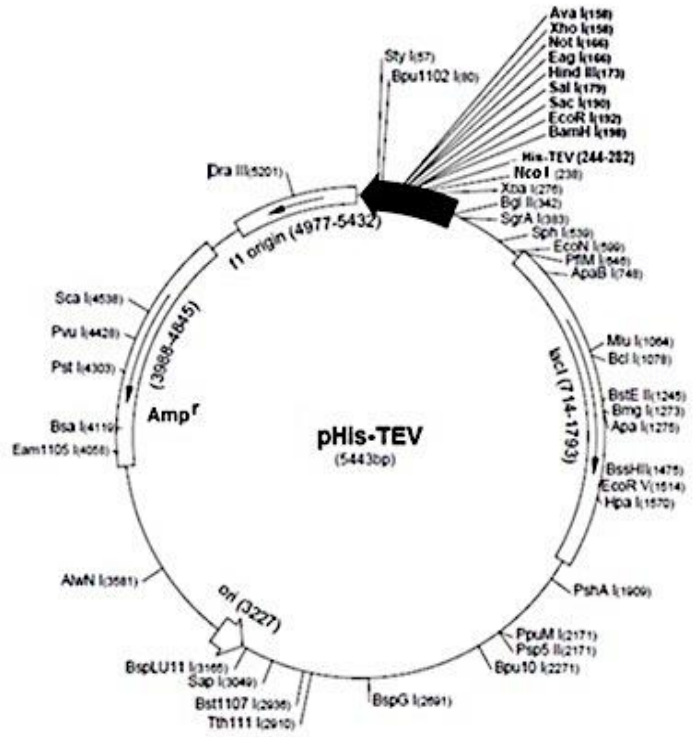


Figure 8.2. Map of pHisTEV vector

8.2. Recipe of media for bacterial culture

Table 8.1. Recipe of LB media

Component	g/L
Trypton	10
Yeast extract	5
NaCl	10

Table 8.2. Recipe of TB media

Component	g/L
Trypton	12
Yeast extract	24
$(NH_4)SO_4$	3.3
KH_2PO_4	6.8
Na_2HPO_4	7.1
Glucose	0.5
Lactose	2
$MgSO_4$	0.15

Table 8.3. Recipe of SOC media

Fomular	g/L
Trypton	20
Yeast extract	*5*
NaCl	0.584
KCl	*0.186*
$MgCl_2$	0.952
$MgSO_4$	*2.467*
glucose	3.603

8.3. Recipe of SDS-PAGE Gels and buffers

Table 8.4. Recipe of SDS-PAGE gels

Component	**10% running gel**	**12% running gel**	**4% Stacking gel**
	Volume		
40% Acrylamide	2.0 ml	2.4 ml	0.5 ml
ddH_2O	3.8 ml	3.4 ml	3.1 ml
1.5M Tris, pH 8.8	2.0 ml	2.0 ml	-
0.5M Tris, pH 6.8	-	-	1.25 ml
TEMED	8 µl	8 µl	5 µl
10% Aps	80 µl	80 µl	50 µl

Table 8.5. Recipe of 2× SDS-PAGE sample buffer

Component	**Volumn (ml)**
1M Tris, pH 6.8	2
50% glycerol	4. 6
10% SDS	1. 6
0.5% bromophenol blue	0. 4
β-mercaptoethanol	0. 4

Table 8.6. Recipe of 10×Tris-Glycine running buffer

Component	weight (g)
Tris base	30
glycine	144
SDS	10
Dissolve the components above in 1000ml of ddH_2O, no pH adjustment is required	

Table 8.7. Recipe of coomassie solution for SDS-PAGE staining

Component	weight/volume
Coomassie blue R-250	0.25 g
EtOH	100 ml
ddH_2O	100 ml

Table 8.8. Recipe of destain solution for SDS-PAGE

Component	volume (ml)
HoAc	100
EtOH	400
ddH_2O	500

8.4. Recipes of TcTS/sialidase activity buffer

TcTS kinetic assay buffer: 20mM Tris, 50mM NaCl, 0.01% Triton X100, pH7.5

CpNA/SpNA kinetic assay buffer: 100mM NaOAc, 2mM $CaCl_2$, 0.01% Triton X100, pH5.0.

8.5. DNA sequences

TcTS_7Mu DNA sequence

>TcTS_7Mu

AAAATAAGGAGAAACATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCCTGGCACCCG
GATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAAGCTCGAAGGTGCCATTTGAAAAGGACGGCAAAG
TCACCGAGCGGGTTGTCCACTCGTTCCGCCTCCCCGCCCTTGTTAATGTGGACGGGGTGATGGTTGCC
ATCGCGGACGCTCGCTACGAAACATCCTTTGACAACTCCCTCATTGATACGGTGGCGAAGTACAGCG
TGGACGATGGGGAGACGTGGGAGACCCAAATTGCCATCAAGAACAGTCGTGCATCGTCTGTTTCTCG
TGTGGTGGATCCCACAGTGATTGTGAAGGGCAACAAGCTTTACGTCCTGGTTGGAAGCTACAACAGT
TCGAGGAGCTACTGGACGTCGCATGGTGATGCGAGAGACTGGGATATTCTGCTTGCCGTTGGTGAGG
TCACGAAGTCCACTGCGGGCGGCAAGATAACTGCGAGTATCAAATGGGGGAGCCCCGTGTCACTGA
AGGAATTTTTCCCGGCGGAAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCAGGTGTTGC
CATTGTGGCGTCCAACGGGAATCTTGTGTACCCTGTGCAGGTTACGAACAAAAAGAAGCAAGTTTTT
TCCAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAGTTTGGGAAGGGTAGGAGCGCTTTTG
GCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATCATAAACACTCGAGTTGACTATCG
CCGCCGTCTGGTGTACGAGTCCAGTGACATGGGGAATACGTGGCTGGAGGCTGTCGGCACGCTCTCA
CGTGTGTGGGGCCCCTCACCAAAATCGAACCAGCCCGGCAGTCAGAGCAGCTTCACTGCCGTGACCA
TCGAGGGAATGCGTGTTATGCTCTTCACACACCCGCTGAATTTTAAGGGAAGGTGGCTGCGCGACCG
ACTGAACCTCTGGCTGACGGATAACCAGCGCATTTATAACGTTGGGCAAGTATCCATTGGTGATGAA
AATTCCGCCTACAGCTCCGTCCTGTACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA
ACGAGGTGTACAGCCTTGTTTTTGCGCGCCTGGTTGGCGAGCTACGGATCATTAAATCAGTGCTGCA
GTCCTGGAAGAATTGGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATCCAGCCGCTTCGTCGT
CAGAGCGTGGTTGTGGTCCCGCTGTCACCACGGTTGGTCTTGTTGGCTTTTTGTCGCACAGTGCCACC
AAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACGCAAGCACGGCAAATGCGGAGAGGGTTCCG
AACGGTTTGAAGTTTGCGGGGGTTGGCGGAGGGGCGCTTTGGCCGGTGAGCCAGCAGGGGCAGAAT
CAACGGTATCACTTTGCAAACCACGCGTTCACGCTGGTGGCGTCGGTGACGATTCACGAGGTTCCGA
AAGGCGCGAGTCCTTTGCTGGGTGCGAGCCTGGACTCTTCTGGTGGCAAAAAACTCCTGGGGCTCTC
GTACGACAAGAGGCACCAGTGGCAGCCAATATACGGATCAACGCCGGTGACGCCGACCGGATCGTG
GGAGATGGGTAAGAGGTACCACGTGGTTCTTACGATGGCGAATAAAATTGGCTCCGTGTACATTGAT
GGAGAACCTCTGGAGGGTTCAGGGCAGACCGTTGTGCCAGACGAGAGGACGCCTGACATCTCCCACT
TCTACGTTGGCGGGTATAAAAGGAGTGGTATGCCAACCGATAGCCGCGTGACGGTGAATAATGTTCT

TC TTTACAACCGTCAGCTGAATGCCGAGGAGATCAGGACCTTGTTCTTGAGCCAGGACCTGATTGGC
ACGGAAGCACACATGGACTGAGCTAGCGAATTCGAAGCTGGCTGTTGCGAAAAAAATTG

TcTS_R249A DNA sequence

AAAATAAGGAGAAACATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCCTGGCACCCG
GATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAAGCTCGAAGGTGCCATTTGAAAAGGACGGCAAAG
TCACCGAGCGGGTTGTCCACTCGTTCCGCCTCCCCGCCCTTGTTAATGTGGACGGGGTGATGGTTGCC
ATCGCGGACGCTCGCTACGAAACATCCTTTGACAACTCCCTCATTGATACGGTGGCGAAGTACAGCG
TGGACGATGGGGAGACGTGGGAGACCCAAATTGCCATCAAGAACAGTCGTGCATCGTCTGTTTCTCG
TGTGGTGGATCCCACAGTGATTGTGAAGGGCAACAAGCTTTACGTCCTGGTTGGAAGCTACAACAGT
TCGAGGAGCTACTGGACGTCGCATGGTGATGCGAGAGACTGGGATATTCTGCTTGCCGTTGGTGAGG
TCACGAAGTCCACTGCGGGCGGCAAGATAACTGCGAGTATCAAATGGGGGAGCCCCGTGTCACTGA
AGGAATTTTTCCCGGCGGAAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCAGGTGTTGC
CATTGTGGCGTCCAACGGGAATCTTGTGTACCCTGTGCAGGTTACGAACAAAAAGAAGCAAGTTTTT
TCCAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAGTTTGGGAAGGGTAGGAGCGCTTTTG
GCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATCATAAACACTCGAGTTGACTATGC
CCGCCGTCTGGTGTACGAGTCCAGTGACATGGGGAATACGTGGCTGGAGGCTGTCGGCACGCTCTCA
CGTGTGTGGGGCCCCTCACCAAAATCGAACCAGCCCGGCAGTCAGAGCAGCTTCACTGCCGTGACCA
TCGAGGGAATGCGTGTTATGCTCTTCACACACCCGCTGAATTTTAAGGGAAGGTGGCTGCGCGACCG
ACTGAACCTCTGGCTGACGGATAACCAGCGCATTTATAACGTTGGGCAAGTATCCATTGGTGATGAA
AATTCCGCCTACAGCTCCGTCCTGTACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA
ACGAGGTGTACAGCCTTGTTTTTGCGCGCCTGGTTGGCGAGCTACGGATCATTAAATCAGTGCTGCA
GTCCTGGAAGAATTGGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATCCAGCCGCTTCGTCGT
CAGAGCGTGGTTGTGGTCCCGCTGTCACCACGGTTGGTCTTGTTGGCTTTTTGTCGCACAGTGCCACC
AAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACGCAAGCACGGCAAATGCGGAGAGGGTTCCG
AACGGTTTGAAGTTTGCGGGGGTTGGCGGAGGGGCGCTTTGGCCGGTGAGCCAGCAGGGGCAGAAT
CAACGGTATCACTTTGCAAACCACGCGTTCACGCTGGTGGCGTCGGTGACGATTCACGAGGTTCCGA
AAGGCGCGAGTCCTTTGCTGGGTGCGAGCCTGGACTCTTCTGGTGGCAAAAAACTCCTGGGGCTCTC
GTACGACAAGAGGCACCAGTGGCAGCCAATATACGGATCAACGCCGGTGACGCCGACCGGATCGTG
GGAGATGGGTAAGAGGTACCACGTGGTTCTTACGATGGCGAATAAAATTGGCTCCGTGTACATTGAT
GGAGAACCTCTGGAGGGTTCAGGGCAGACCGTTGTGCCAGACGAGAGGACGCCTGACATCTCCCACT
TCTACGTTGGCGGGTATAAAAGGAGTGGTATGCCAACCGATAGCCGCGTGACGGTGAATAATGTTCT
TC TTTACAACCGTCAGCTGAATGCCGAGGAGATCAGGACCTTGTTCTTGAGCCAGGACCTGATTGGC
ACGGAAGCACACATGGACTGAGCTAGCGAATTCGAAGCTGGCTGTTGCGAAAAAAATTG

TcTS_R272A DNA sequence

AAAATAAGGAGAAACATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCCTGGCACCCG
GATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAAGCTCGAAGGTGCCATTTGAAAAGGACGGCAAAG
TCACCGAGCGGGTTGTCCACTCGTTCCGCCTCCCCGCCCTTGTTAATGTGGACGGGGTGATGGTTGCC
ATCGCGGACGCTCGCTACGAAACATCCTTTGACAACTCCCTCATTGATACGGTGGCGAAGTACAGCG

TGGACGATGGGGAGACGTGGGAGACCCAAATTGCCATCAAGAACAGTCGTGCATCGTCTGTTTCTCG
TGTGGTGGATCCCACAGTGATTGTGAAGGGCAACAAGCTTTACGTCCTGGTTGGAAGCTACAACAGT
TCGAGGAGCTACTGGACGTCGCATGGTGATGCGAGAGACTGGGATATTCTGCTTGCCGTTGGTGAGG
TCACGAAGTCCACTGCGGGCGGCAAGATAACTGCGAGTATCAAATGGGGGAGCCCCGTGTCACTGA
AGGAATTTTTCCCGGCGGAAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCAGGTGTTGC
CATTGTGGCGTCCAACGGGAATCTTGTGTACCCTGTGCAGGTTACGAACAAAAAGAAGCAAGTTTTT
TCCAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAGTTTGGGAAGGGTAGGAGCGCTTTTG
GCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATCATAAACACTCGAGTTGACTATCG
CCGCCGTCTGGTGTACGAGTCCAGTGACATGGGGAATACGTGGCTGGAGGCTGTCGGCACGCTCTCA
GCTGTGTGGGGCCCCTCACCAAAATCGAACCAGCCCGGCAGTCAGAGCAGCTTCACTGCCGTGACCA
TCGAGGGAATGCGTGTTATGCTCTTCACACACCCGCTGAATTTTAAGGGAAGGTGGCTGCGCGACCG
ACTGAACCTCTGGCTGACGGATAACCAGCGCATTTATAACGTTGGGCAAGTATCCATTGGTGATGAA
AATTCCGCCTACAGCTCCGTCCTGTACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA
ACGAGGTGTACAGCCTTGTTTTTGCGCGCCTGGTTGGCGAGCTACGGATCATTAAATCAGTGCTGCA
GTCCTGGAAGAATTGGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATCCAGCCGCTTCGTCGT
CAGAGCGTGGTTGTGGTCCCGCTGTCACCACGGTTGGTCTTGTTGGCTTTTTGTCGCACAGTGCCACC
AAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACGCAAGCACGGCAAATGCGGAGAGGGTTCCG
AACGGTTTGAAGTTTGCGGGGGTTGGCGGAGGGGCGCTTTGGCCGGTGAGCCAGCAGGGGCAGAAT
CAACGGTATCACTTTGCAAACCACGCGTTCACGCTGGTGGCGTCGGTGACGATTCACGAGGTTCCGA
AAGGCGCGAGTCCTTTGCTGGGTGCGAGCCTGGACTCTTCTGGTGGCAAAAAACTCCTGGGGCTCTC
GTACGACAAGAGGCACCAGTGGCAGCCAATATACGGATCAACGCCGGTGACGCCGACCGGATCGTG
GGAGATGGGTAAGAGGTACCACGTGGTTCTTACGATGGCGAATAAAATTGGCTCCGTGTACATTGAT
GGAGAACCTCTGGAGGGTTCAGGGCAGACCGTTGTGCCAGACGAGAGGACGCCTGACATCTCCCACT
TCTACGTTGGCGGGTATAAAAGGAGTGGTATGCCAACCGATAGCCGCGTGACGGTGAATAATGTTCT
TC TTTACAACCGTCAGCTGAATGCCGAGGAGATCAGGACCTTGTTCTTGAGCCAGGACCTGATTGGC
ACGGAAGCACACATGGACTGAGCTAGCGAATTCGAAGCTGGCTGTTGCGAAAAAAATTG

TcTS_R577A DNA sequence

AAAATAAGGAGAAACATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCCTGGCACCCG
GATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAAGCTCGAAGGTGCCATTTGAAAAGGACGGCAAAG
TCACCGAGCGGGTTGTCCACTCGTTCCGCCTCCCCGCCCTTGTTAATGTGGACGGGGTGATGGTTGCC
ATCGCGGACGCTCGCTACGAAACATCCTTTGACAACTCCCTCATTGATACGGTGGCGAAGTACAGCG
TGGACGATGGGGAGACGTGGGAGACCCAAATTGCCATCAAGAACAGTCGTGCATCGTCTGTTTCTCG
TGTGGTGGATCCCACAGTGATTGTGAAGGGCAACAAGCTTTACGTCCTGGTTGGAAGCTACAACAGT
TCGAGGAGCTACTGGACGTCGCATGGTGATGCGAGAGACTGGGATATTCTGCTTGCCGTTGGTGAGG
TCACGAAGTCCACTGCGGGCGGCAAGATAACTGCGAGTATCAAATGGGGGAGCCCCGTGTCACTGA
AGGAATTTTTCCCGGCGGAAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCAGGTGTTGC
CATTGTGGCGTCCAACGGGAATCTTGTGTACCCTGTGCAGGTTACGAACAAAAAGAAGCAAGTTTTT
TCCAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAGTTTGGGAAGGGTAGGAGCGCTTTTG
GCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATCATAAACACTCGAGTTGACTATCG
CCGCCGTCTGGTGTACGAGTCCAGTGACATGGGGAATACGTGGCTGGAGGCTGTCGGCACGCTCTCA

CGTGTGTGGGGCCCCTCACCAAAATCGAACCAGCCCGGCAGTCAGAGCAGCTTCACTGCCGTGACCA
TCGAGGGAATGCGTGTTATGCTCTTCACACACCCGCTGAATTTTAAGGGAAGGTGGCTGCGCGACCG
ACTGAACCTCTGGCTGACGGATAACCAGCGCATTTATAACGTTGGGCAAGTATCCATTGGTGATGAA
AATTCCGCCTACAGCTCCGTCCTGTACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA
ACGAGGTGTACAGCCTTGTTTTTGCGCGCCTGGTTGGCGAGCTACGGATCATTAAATCAGTGCTGCA
GTCCTGGAAGAATTGGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATCCAGCCGCTTCGTCGT
CAGAGCGTGGTTGTGGTCCCGCTGTCACCACGGTTGGTCTTGTTGGCTTTTTGTCGCACAGTGCCACC
AAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACGCAAGCACGGCAAATGCGGAGAGGGTTCCG
AACGGTTTGAAGTTTGCGGGGGTTGGCGGAGGGGCGCTTTGGCCGGTGAGCCAGCAGGGGCAGAAT
CAACGGTATCACTTTGCAAACCACGCGTTCACGCTGGTGGCGTCGGTGACGATTCACGAGGTTCCGA
AAGGCGCGAGTCCTTTGCTGGGTGCGAGCCTGGACTCTTCTGGTGGCAAAAAACTCCTGGGGCTCTC
GTACGACAAGAGGCACCAGTGGCAGCCAATATACGGATCAACGCCGGTGACGCCGACCGGATCGTG
GGAGATGGGTAAGAGGTACCACGTGGTTCTTACGATGGCGAATAAAATTGGCTCCGTGTACATTGAT
GGAGAACCTCTGGAGGGTTCAGGGCAGACCGTTGTGCCAGACGAGGCGACGCCTGACATCTCCCACT
TCTACGTTGGCGGGTATAAAAGGAGTGGTATGCCAACCGATAGCCGCGTGACGGTGAATAATGTTCT
TC TTTACAACCGTCAGCTGAATGCCGAGGAGATCAGGACCTTGTTCTTGAGCCAGGACCTGATTGGC
ACGGAAGCACACATGGACTGAGCTAGCGAATTCGAAGCTGGCTGTTGCGAAAAAAATTG

<u>TcTS_D59N DNA sequence</u>

AAAATAAGGAGAAACATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCCTGGCACCCG
GATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAAGCTCGAAGGTGCCATTTGAAAAGGACGGCAAAG
TCACCGAGCGGGTTGTCCACTCGTTCCGCCTCCCCGCCCTTGTTAATGTGGACGGGGTGATGGTTGCC
ATCGCGGACGCTCGCTACGAAACATCCTTTAACAACTCCCTCATTGATACGGTGGCGAAGTACAGCG
TGGACGATGGGGAGACGTGGGAGACCCAAATTGCCATCAAGAACAGTCGTGCATCGTCTGTTTCTCG
TGTGGTGGATCCCACAGTGATTGTGAAGGGCAACAAGCTTTACGTCCTGGTTGGAAGCTACAACAGT
TCGAGGAGCTACTGGACGTCGCATGGTGATGCGAGAGACTGGGATATTCTGCTTGCCGTTGGTGAGG
TCACGAAGTCCACTGCGGGCGGCAAGATAACTGCGAGTATCAAATGGGGGAGCCCCGTGTCACTGA
AGGAATTTTTCCCGGCGGAAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCAGGTGTTGC
CATTGTGGCGTCCAACGGGAATCTTGTGTACCCTGTGCAGGTTACGAACAAAAAGAAGCAAGTTTTT
TCCAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTGGAAGTTTGGGAAGGGTAGGAGCGCTTTTG
GCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATCATAAACACTCGAGTTGACTATCG
CCGCCGTCTGGTGTACGAGTCCAGTGACATGGGGAATACGTGGCTGGAGGCTGTCGGCACGCTCTCA
CGTGTGTGGGGCCCCTCACCAAAATCGAACCAGCCCGGCAGTCAGAGCAGCTTCACTGCCGTGACCA
TCGAGGGAATGCGTGTTATGCTCTTCACACACCCGCTGAATTTTAAGGGAAGGTGGCTGCGCGACCG
ACTGAACCTCTGGCTGACGGATAACCAGCGCATTTATAACGTTGGGCAAGTATCCATTGGTGATGAA
AATTCCGCCTACAGCTCCGTCCTGTACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA
ACGAGGTGTACAGCCTTGTTTTTGCGCGCCTGGTTGGCGAGCTACGGATCATTAAATCAGTGCTGCA
GTCCTGGAAGAATTGGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATCCAGCCGCTTCGTCGT
CAGAGCGTGGTTGTGGTCCCGCTGTCACCACGGTTGGTCTTGTTGGCTTTTTGTCGCACAGTGCCACC
AAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACGCAAGCACGGCAAATGCGGAGAGGGTTCCG
AACGGTTTGAAGTTTGCGGGGGTTGGCGGAGGGGCGCTTTGGCCGGTGAGCCAGCAGGGGCAGAAT

CAACGGTATCACTTTGCAAACCACGCGTTCACGCTGGTGGCGTCGGTGACGATTCACGAGGTTCCGA
AAGGCGCGAGTCCTTTGCTGGGTGCGAGCCTGGACTCTTCTGGTGGCAAAAAACTCCTGGGGCTCTC
GTACGACAAGAGGCACCAGTGGCAGCCAATATACGGATCAACGCCGGTGACGCCGACCGGATCGTG
GGAGATGGGTAAGAGGTACCACGTGGTTCTTACGATGGCGAATAAAATTGGCTCCGTGTACATTGAT
GGAGAACCTCTGGAGGGTTCAGGGCAGACCGTTGTGCCAGACGAGGCGACGCCTGACATCTCCCACT
TCTACGTTGGCGGGTATAAAAGGAGTGGTATGCCAACCGATAGCCGCGTGACGGTGAATAATGTTCT
TC TTTACAACCGTCAGCTGAATGCCGAGGAGATCAGGACCTTGTTCTTGAGCCAGGACCTGATTGGC
ACGGAAGCACACATGGACTGAGCTAGCGAATTCGAAGCTGGCTGTTGCGAAAAAAATTG

8.6. Interaction details of TcTS and inhibitors

Table 8.9. Details of interaction between TcTS and naphthalene compounds

Residue No.	Amino acid	Compound component	Symmetry related molecule	Domain	Distance (Å)	Mode of Interaction
Crystal 1, compound 51, P1 2_1 1						
540	Glu	Naphthalene	No	LT*	3.6	Hydrophobic interaction
543	Lys	Naphthalene	No	LT	3.5	Hydrophobic interaction
544	Arg	Phosphate oxygen	No	LT	2.71	Hydrogen bond
90	Ser	Phosphate oxygen	Yes	CA	3.33	Hydrogen bond
Crystal 2, compound 54, P1 2_1 1						
540	Glu	Naphthalene	No	LT	3.79	Hydrophobic interaction
543	Lys	Naphthalene	No	LT	3.49	Hydrophobic interaction
544	Arg	Phosphate oxygen	No	LT	2.73	Hydrogen bond
90	Ser	Phosphate oxygen	Yes	CA*	3.3	Hydrogen bond
Crystal 3, compound 54, P 2_1 2_1 2_1						
522	His	Naphthalene	No	LT	3.77	Hydrophobic interaction
520	Glu	Naphthalene	No	LT	3.75	Hydrophobic interaction
538	Ser	Phosphate oxygen	No	LT	2.15	Hydrogen bond
530	Ser	Phosphate oxygen	Yes	LT	2.32	Hydrogen bond
579	Pro	Nitro oxygen	Yes	LT	3.5	Salt bridge

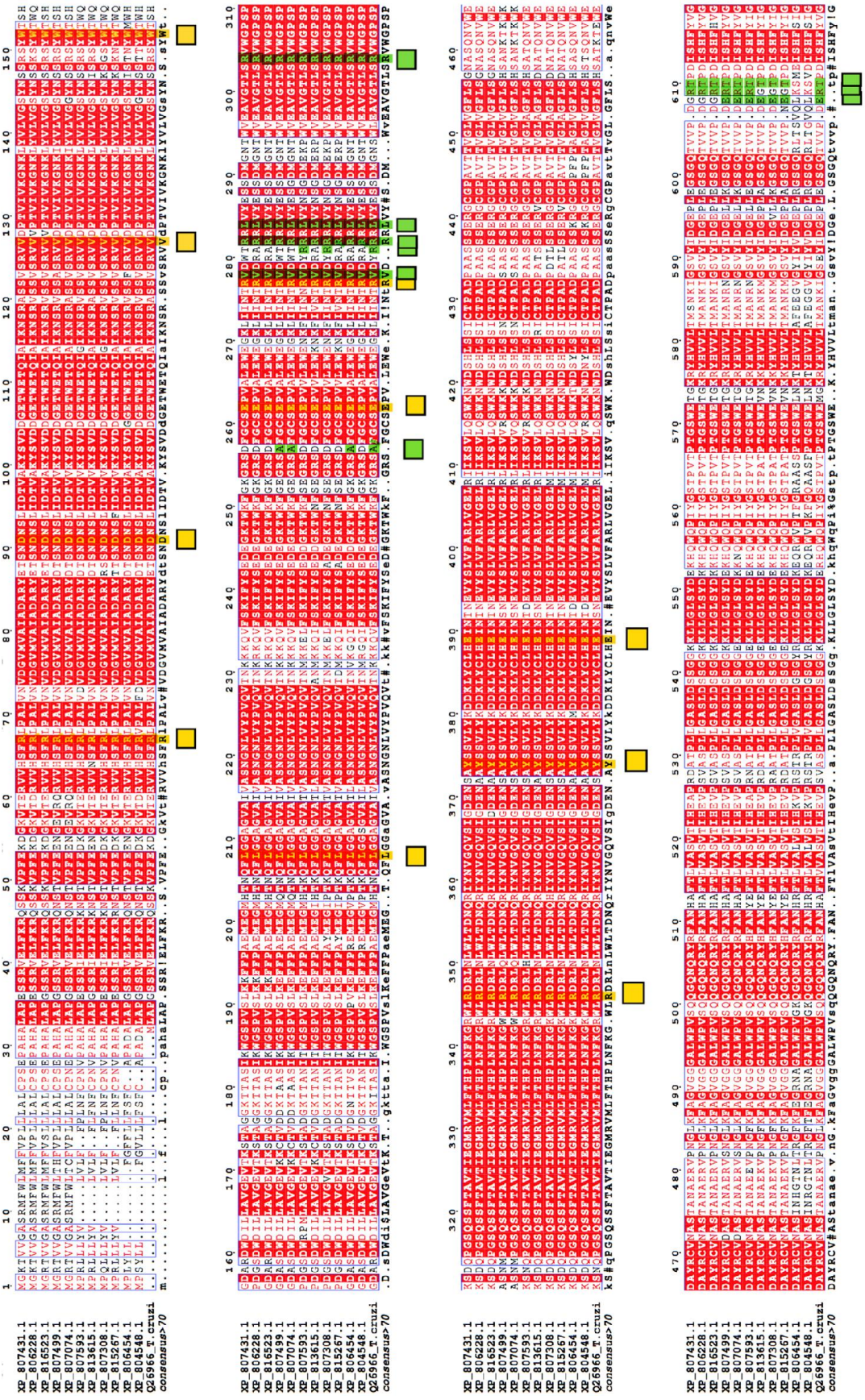
* CA: catalytic domain； LT: lectin-like domain

Table 8.10. Details of interaction between TcTS and anthracene compound **48**

Residue No.	AA	Compound component	Domain	Distance (Å)	Mode of Interaction
Crystal 4, compound 48, P 2_1 2_1 2_1, Hilex binding site					
376	Arg	Anthracene	CA	3.6-3.8	Hydrophobic interaction
376	Arg	Amino nitrogen	CA	3.8	Hydrogen bond
376	Arg	Anthracene	CA	3.5	π-Cation interaction
377	Ile	Anthracene	CA	3.6-3.7	Hydrophobic interaction
380	Ser	Phosphate oxygen	CA	2.52	Hydrogen bond
437	Arg	Phosphate oxygen	CA	3	Salt bridge
Crystal 4, compound 48, P 2_1 2_1 2_1, Inter-domains binding site					
225	Ala	Anthracene	CA	3.71	Hydrophobic interaction
246	Val	Anthracene	CA	3.8	Hydrophobic interaction
249	Arg	Anthracene	CA	3.9	Hydrophobic interaction
249*	Arg	Anthracene	CA	3.7	π-Cation interaction
249	Arg	Anthracene	CA	3.8	π-Cation interaction
250	Arg	Anthracene	CA	4	Hydrophobic interaction
252	Leu	Anthracene	CA	3.5	Hydrophobic interaction
272	Arg	Anthracene	CA	3.9	Hydrophobic interaction
272	Arg	Phosphate oxygen	CA	3.2	Salt bridge
577	Arg	Phosphate oxygen	LT	3.1	Salt bridge
577	Arg	Phenyl	LT	3.6	π-Cation interaction
578	Thr	Phosphate oxygen	LT	3.4	Hydrogen bond

*The two π-cation interactions of Arg249 are interact with different aromatic core of anthracene.

8.7. Sequence alignment of TcTS



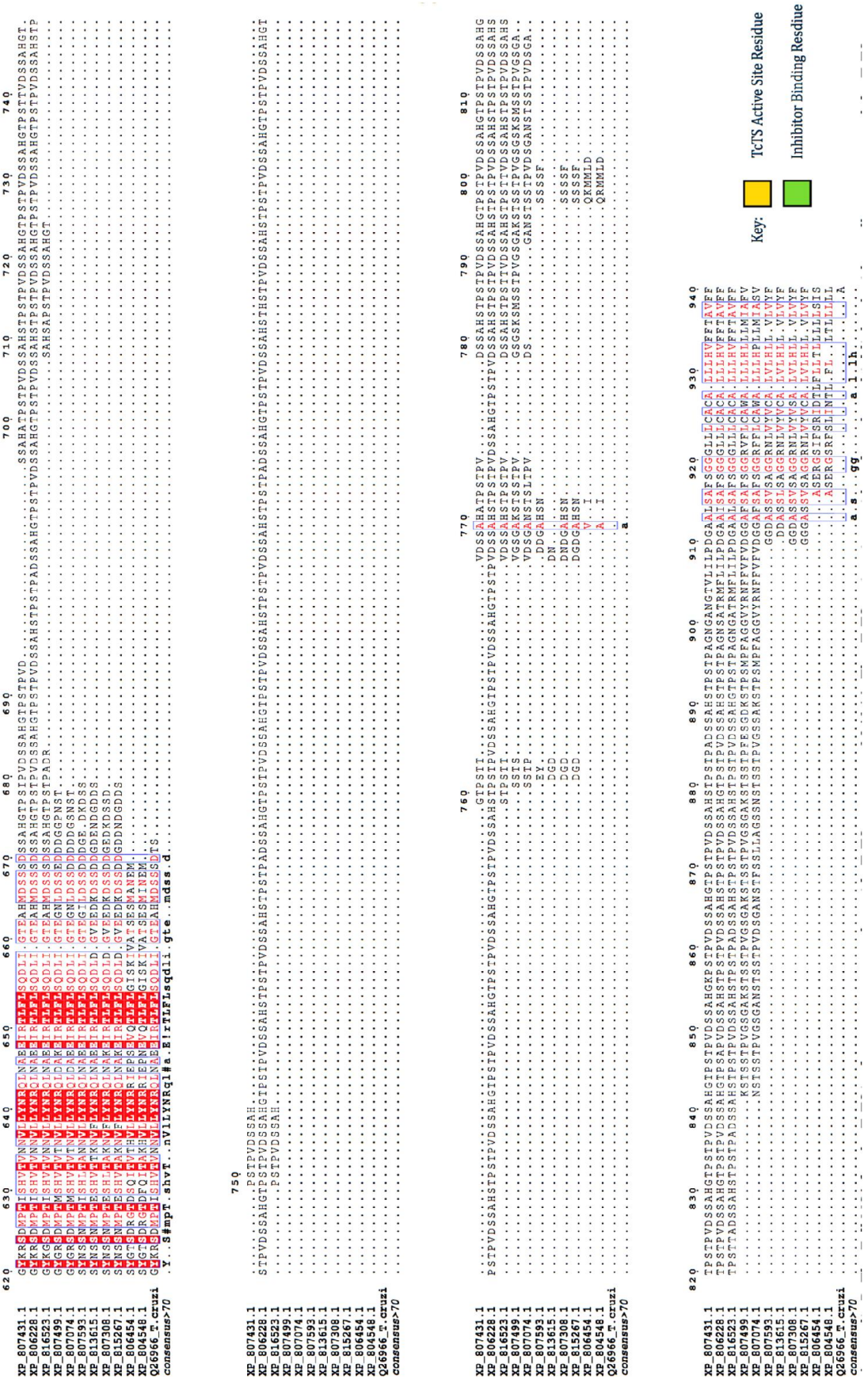


Figure 8.8.1. Full multiple sequence alignment result of TcTS sequences of CL banner strains alongside sequence Q26966. The catalytic site residues are highlighted with yellow box and the

allosteric binding site residues are highlighted by green box under the alignment results. The conversation of these residues within TcTS isoform has been further highlighted in yellow (catalytic site) and green (allosteric binding site).

Chapter 9. Bibliography

Acharya, K. R., and Lloyd, M. D. (2005). The advantages and limitations of protein crystal structures. *Trends Pharmacol Sci, 26*, 10-14.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M. and Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr, 68*, 352-367.

Agusti, R., Cano, M. E., Cagnoni, A. J., Kovensky, J., de Lederkremer, R. M., and Uhrig, M. L. (2016). Multivalent sialylation of beta-thio-glycoclusters by *Trypanosoma cruzi* trans sialidase and analysis by high performance anion exchange chromatography. *Glycoconjugate Journal, 33*, 809-818

Agusti, R., Couto, A. S., Alves, M. J. M., Colli, W., and de Lederkremer, R. M. (2000). Lipids shed into the culture medium by trypomastigotes of *Trypanosoma cruzi*. *Memorias Do Instituto Oswaldo Cruz, 95*, 97-102.

Agusti, R., Couto, A. S., Campetella, O. E., Frasch, A. C. C., and deLederkremer, R. M. (1997). The trans-sialidase of *Trypanosoma cruzi* is anchored by two different lipids. *Glycobiology, 7*, 731-735.

Agusti, R., Giorgi, M. E., Mendoza, V. M., Gallo-Rodriguez, C., and de Lederkremer, R. M. (2007). Comparative rates of sialylation by recombinant trans-sialidase and inhibitor properties of synthetic oligosaccharides from *Trypanosoma cruzi* mucins-containing galactofuranose and galactopyranose. *Bioorg Med Chem, 15*, 2611-2616.

Agusti, R., Paris, G., Ratier, L., Frasch, A. C., and de Lederkremer, R. M. (2004). Lactose derivatives are inhibitors of *Trypanosoma cruzi* trans-sialidase activity toward conventional substrates in vitro and in vivo. *Glycobiology, 14*, 659-670.

Amaya, M. F., Watts, A. G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A. and Alzari, P. M. (2004). Structural insights into the catalytic mechanism of Trypanosoma cruzi trans-sialidase. *Structure, 12*, 775-784.

Amino, R., Porto, R. M., Chammas, R., Egami, M. I., and Schenkman, S. (1998). Identification and characterization of a sialidase released by the salivary gland of the hematophagous insect Triatoma infestans. *J Biol Chem, 273*, 24575-24582.

Angata, T., and Varki, A. (2002). Chemical diversity in the sialic acids and related alpha-keto

acids: an evolutionary perspective. *Chem Rev, 102*, 439-469.

Arioka, S., Sakagami, M., Uematsu, R., Yamaguchi, H., Togame, H., Takemoto, H. and Nishimura, S. (2010). Potent inhibitor scaffold against Trypanosoma cruzi trans-sialidase. *Bioorg Med Chem, 18*, 1633-1640.

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B. and Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res, 34*, 604-608.

Arnaud, J., Audfray, A., and Imberty, A. (2013). Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. *Chem Soc Rev, 42*, 4798-4813.

Atwood, J. A., 3rd, Weatherly, D. B., Minning, T. A., Bundy, B., Cavola, C. and Opperdoes, F. R. and Tarleton, R. L. (2005). The Trypanosoma cruzi proteome. *Science, 309*, 473-476.

Bai, X., McMullan, G., and Scheres, S. H.. (2015). How cryo-EM is revolutionizing structural biology. *Trends in biochemical sciences, 40*, 49-57.

Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr, 67*, 271-281.

Bern, C., Montgomery, S. P., Herwaldt, B. L., Rassi, A., Jr., Marin-Neto, J. A., Dantas, R. O. and Moore, A. C. (2007). Evaluation and treatment of chagas disease in the United States: a systematic review. *JAMA, 298*, 2171-2181.

Blanchet, C. E., and Svergun, D. I. (2013). Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu Rev Phys Chem, 64*, 37-54.

Bolognesi, C., Parrini, M., Aiello, C., and Rossi, L. (1991). DNA damage induced by 7,12-dimethylbenz[a]anthracene in the liver and the mammary gland of rats exposed to polycyclic aromatic hydrocarbon enzyme inducers during perinatal life. *Mutagenesis, 6*(2), 113-116.

Bragg, W. H., and Bragg, W. L. (1913). The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 88*(605), 428-438.

Brunger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature, 355*, 472-475.

Brunger, A. T., and Rice, L. M. (1997). Crystallographic refinement by simulated annealing: methods and applications. *Methods Enzymol, 277*, 243-269.

Buchini, S., Buschiazzo, A., and Withers, S. G. (2008). A new generation of specific Trypanosoma cruzi trans-sialidase inhibitors. *Angew Chem Int Ed Engl, 47*, 2700-2703.

Bunkoczi, G., Echols, N., McCoy, A. J., Oeffner, R. D., Adams, P. D., and Read, R. J. (2013). Phaser.MRage: automated molecular replacement. *Acta Crystallogr D Biol Crystallogr, 69*, 2276-2286.

Buscaglia, C. A., Alfonso, J., Campetella, O., and Frasch, A. C. C. (1999). Tandem amino acid repeats from Trypanosoma cruzi shed antigens increase the half-life of proteins in blood. *Blood, 93*, 2025-2032.

Buscaglia, C. A., Campetella, O., Leguizamon, M. S., and Frasch, A. C. C. (1998). The repetitive domain of Trypanosoma cruzi trans-sialidase enhances the immune response against the catalytic domain. *Journal of Infectious Diseases, 177*, 431-436.

Buscaglia, C. A., Campo, V. A., Frasch, A. C., and Di Noia, J. M. (2006). Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nat Rev Microbiol, 4*, 229-236.

Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C., and Alzari, P. M. (2002). The crystal structure and mode of action of trans-sialidase, a key enzyme in Trypanosoma cruzi pathogenesis. *Mol Cell, 10*, 757-768.

Buschiazzo, A., Tavares, G. A., Campetella, O., Spinelli, S., Cremona, M. L., Paris, G. and Alzari, P. M. (2000). Structural basis of sialyltransferase activity in trypanosomal sialidases. *EMBO J, 19*, 16-24.

Busse, H., Hakoda, M., Stanley, M., and Streicher, H. (2007). Galactose-phosphonates as mimetics of the sialyltransfer by trypanosomal sialidases. *Journal of Carbohydrate Chemistry, 26*, 159-194.

Caboche, S. (2013). LeView: automatic and interactive generation of 2D diagrams for biomacromolecule/ligand interactions. *Journal of Cheminformatics, 5*.

Cano, M. E., Agusti, R., Cagnoni, A. J., Tesoriero, M. F., Kovensky, J., Uhrig, M. L., and de Lederkremer, R. M. (2014). Synthesis of divalent ligands of beta-thio- and beta-N-galacto-pyranosides and related lactosides and their evaluation as substrates and inhibitors of Trypanosoma cruzi trans-sialidase. *Beilstein Journal of Organic Chemistry, 10*, 3073-3086.

Carvalho, I., Andrade, P., Campo, V. L., Guedes, P. M., Sesti-Costa, R., Silva, J. S. and Field, R. A. (2010). 'Click chemistry' synthesis of a library of 1,2,3-triazole-substituted galactose derivatives and their evaluation against Trypanosoma cruzi and its cell surface trans-sialidase. *Bioorg Med Chem, 18*(7), 2412-2427. doi: 10.1016/j.bmc.2010.02.053

Carvalho ., Sola-Penna, M., Oliveira, I. A., Pita, S., Goncalves, A. S., Neves, B. C., and Todeschini, A. R. (2010). A new class of mechanism-based inhibitors for Trypanosoma cruzi trans-sialidase and their influence on parasite virulence. *Glycobiology, 20*, 1034-1045.

Castro, J. A, Montalto M., María, and Bartel, L. C. (2006). Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Human and experimental toxicology, 25*(8), 471-479.

Cavanagh, J., Fairbrother, W. J., Palmer III, A., and Skelton, N. J. (1995). *Protein NMR spectroscopy: principles and practice*: Academic Press.

CDC. (2014). CDC Parasites - American Trypanosomiasis (also known as Chagas Disease). from http://www.cdc.gov/parasites/chagas/

Cecere, M. C., Castanera, M. B., Canale, D. M., Chuit, R., and Gurtler, R. E. (1999). Trypanosoma cruzi infection in Triatoma infestans and other triatomines: long-term effects of a control program in rural northwestern Argentina. *Rev Panam Salud Publica, 5*(6), 392-399.

Chagas, C. (1909). Nova tripanozomiaze humana: estudos sobre a morfolojia e o ciclo evolutivo do Schizotrypanum cruzi n. gen., n. sp., ajente etiolojico de nova entidade morbida do homem. *Memórias do Instituto Oswaldo Cruz, 1*(2), 159-218.

Chan, T. H., Xin, Y. C., and vonItzstein, M. (1997). Synthesis of phosphonic acid analogues of sialic acids (Neu5Ac and KDN) as potential sialidase inhibitors. *Journal of Organic Chemistry, 62*, 3500-3504.

Chand, P., Babu, Y. S., Bantia, S., Chu, N. M., Cole, L. B., Kotian, P. L., Walsh, G. W. (1997). Design and synthesis of benzoic acid derivatives as influenza neuraminidase inhibitors using structure-based drug design. *Journal of Medicinal Chemistry, 40*, 4030-4052.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr, 66*, 12-21.

Chuenkova, M. V., Furnari, F. B., Cavenee, W. K., and Pereira, M. A. (2001). Trypanosoma cruzi trans sialidase: A potent and specific survival factor for human Schwann cells by means of phosphatidylinositol 3-kinase/Akt signaling. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 9936-9941.

Chuenkova, M. V., and Perrin, M. (2004). Chagas' disease parasite promotes neuron survival and differentiation through TrkA nerve growth factor receptor. *Journal of Neurochemistry, 91*, 385-394.

Clayton, J. (2010). Chagas disease 101. *Nature, 465*(7301), S4-5.

Colli, W. (1993). Trans-sialidase: a unique enzyme activity discovered in the protozoan Trypanosoma cruzi. *FASEB J, 7*(13), 1257-1264.

Cordova, E., Maiolo, E., Corti, M., and Orduna, T. (2010). Neurological manifestations of Chagas' disease. *Neurol Res, 32*, 238-244.

Cosconati, S., Forli, S., Perryman, A. L., Harris, R., Goodsell, D. S., and Olson, A. J. (2010). Virtual Screening with AutoDock: Theory and Practice. *Expert Opin Drug Discov, 5*, 597-607.

Coura, J. R., and Vinas, P. A. (2010). Chagas disease: a new worldwide challenge. *Nature, 465*, S6-7.

Coussens, L. M., and Werb, Z. (2002). Inflammation and cancer. *Nature, 420*, 860-867.

Crennell, S., Garman, E., Laver, G., Vimr, E., and Taylor, G.. (1994). Crystal structure of Vibrio cholerae neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure, 2*, 535-544.

Damager, I., Buchini, S., Amaya, M. F., Buschiazzo, A., Alzari, P., Frasch, A. C. and Withers, S. G. (2008). Kinetic and mechanistic analysis of Trypanosoma cruzi trans-sialidase reveals a classical ping-pong mechanism with acid/base catalysis. *Biochemistry, 47*, 3507-3512.

Damian, L. (2013). Isothermal titration calorimetry for studying protein-ligand interactions. *Methods Mol Biol, 1008*, 103-118.

de La Fortelle, E., and Bricogne, G.. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods in enzymology, 276*, 472-494.

Dias, W. B., Fajardo, F. D., Graca-Souza, A. V., Freire-de-Lima, L., Vieira, F., Girard, M. F.,

Todeschini, A. R. (2008). Endothelial cell signalling induced by trans-sialidase from Trypanosoma cruzi. *Cell Microbiol, 10*, 88-99.

Duff, M. R., Jr., Grubbs, J., and Howell, E. E. (2011). Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *J Vis Exp*(55).

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2126-2132.

Enders, D., Thomas, C. R., Vignola, N., and Raabe, G. (2002). Asymmetric synthesis of alpha-substituted N-methylsulfonamides. *Helvetica Chimica Acta, 85*, 3657-3677.

Engstler, M., Ferrerogarcia, M. A., Parodi, A. J., Schauer, R., Storzeckerlin, T., Vasella, A., Zhu, X. Y. (1994). N-(4-Nitrophenyl)Oxamic Acid and Related N-Acylanilines Are Noncompetitive Inhibitors of Vibrio-Cholerae Sialidase but Do Not Inhibit Trypanosoma-Cruzi or Trypanosoma-Brucei Trans-Sialidases. *Helvetica Chimica Acta, 77*, 1166-1174.

Feng, B. Y., and Shoichet, B. K. (2006). A detergent-based assay for the detection of promiscuous inhibitors. *Nat Protoc, 1*(2), 550-553.

Ferrerogarcia, M. A., Sanchez, D. O., Frasch, A. C. C., and Parodi, A. J. (1993). The Effect of Pyridoxal 5'-Phosphate and Related-Compounds on Trypanosoma-Cruzi Trans-Sialidase. *Anales De La Asociacion Quimica Argentina, 81*(2-3), 127-132.

Fields, E. K. (1952). The Synthesis of Esters of Substituted Amino Phosphonic Acids. *Journal of the American Chemical Society, 74*(6), 1528-1531.

Flynn, A., Beight, D. W., Bohme, E., and Metcalf, B. W. (1985). The synthesis of fluorinated aminophosphonic acid inhibitors of alanine racemase. *Tetrahedron letters, 26*(3), 285-288.

Freire-De-Lima, L., Alisson-Silva, F., Carvalho ., Takiya, C. M., Rodrigues, M. M., DosReis, G. A., Todeschini, A. R. (2010). Trypanosoma cruzi Subverts Host Cell Sialylation and May Compromise Antigen-specific CD8(+) T Cell Responses. *Journal of Biological Chemistry, 285*, 13388-13396.

Freire-de-Lima, L., da Fonseca, L. M., da Silva, V. A., da Costa, K. M., Morrot, A., Freire-de-Lima, C. G., Mendonca-Previato, L. (2016). Modulation of Cell Sialoglycophenotype: A Stylish Mechanism Adopted by Trypanosoma cruzi to Ensure Its Persistence in the Infected Host. *Frontiers in Microbiology, 7*.

Freire-de-Lima, L., da Fonseca, L. M., da Silva, V. A., da Costa, K. M., Morrot, A., Freire-de-Lima, C. G., Mendonca-Previato, L. (2016). Modulation of Cell Sialoglycophenotype: A Stylish Mechanism Adopted by Trypanosoma cruzi to Ensure Its Persistence in the Infected Host. *Front Microbiol, 7*, 698.

Freire-De-Lima, L., Oliveira, I. A., Neves, J. L., Penha, L. L., Alisson-Silva, F., Dias, W. B., and Todeschini, A. R. (2012). Sialic acid,: a sweet swing between mammalian host and Trypanosoma cruzi. *Frontiers in Immunology, 3*.

Fuster, M. M., and Esko, J. D. (2005). The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nature Reviews Cancer, 5*, 526-542.

Gao, W., Wortis, H. H., and Pereira, M. A. (2002). The Trypanosoma cruzi trans-sialidase is a T cell-independent B cell mitogen and an inducer of non-specific Ig secretion. *International Immunology, 14*, 299-308.

Grienke, U., Schmidtke, M., von Grafenstein, S., Kirchmair, J., Liedl, K. R., and Rollinger, J. M. (2012). Influenza neuraminidase: a druggable target for natural products. *Nat Prod Rep, 29*, 11-36.

Guhl, F., Jaramillo, C., Vallejo, G. A., Cardenas, A. Arroyo F., and Aufderheide, A. (2000). Chagas disease and human migration. *Mem Inst Oswaldo Cruz, 95*, 553-555.

Gutierrez, A.J., Prisbe, E.J., Rohloff, J.C., Methods for the dealkylation of phosphonate esters, US patent US6465649.B1, 1999

Guo, J. H., and Severtson, S. J. (2004). Inhibition of calcium carbonate nucleation with aminophosphonates at high temperature, pH and ionic strength. *Industrial and Engineering Chemistry Research, 43*, 5411-5417.

Hamers, R. L., van Gool, T., and Goorhuis, A. (2016). Benznidazole for Chronic Chagas' Cardiomyopathy. *New England Journal of Medicine, 374*, 188-188.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *J Mol Biol, 166*(4), 557-580.

Harrison, J. A., Kartha, K. P., Fournier, E. J., Lowary, T. L., Malet, C., Nilsson, U. J., Field, R. A. (2011). Probing the acceptor substrate binding site of Trypanosoma cruzi trans-sialidase with systematically modified substrates and glycoside libraries. *Org Biomol Chem, 9*, 1653-1660.

Harrison, J. A., Kartha, K. P., Turnbull, W. B., Scheuerl, S. L., Naismith, J. H., Schenkman, S.,

and Field, R. A. (2001). Hydrolase and sialyltransferase activities of trypanosoma cruzi trans-sialidase towards NeuAc-alpha-2,3-gal-Gal-beta-O-PNP. *Bioorg Med Chem Lett, 11*(2), 141-144.

Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology, 7*(5), 637-644.

Hicks, S. J., Theodoropoulos, G., Carrington, S. D., and Corfield, A. P. (2000). The role of mucins in host-parasite interactions. Part I-protozoan parasites. *Parasitol Today, 16*(11), 476-481.

Hunter, Christopher A, and Sanders, Jeremy KM. (1990). The nature of. pi.-. pi. interactions. *Journal of the American Chemical Society, 112*(14), 5525-5534.

Ilisz, I., Gecse, Z., Pataj, Z., Fulop, F., Toth, G., Lindner, W., and Peter, A. (2014). Direct high-performance liquid chromatographic enantioseparation of secondary amino acids on Cinchona alkaloid-based chiral zwitterionic stationary phases. Unusual temperature behavior. *Journal of Chromatography A, 1363*, 169-177.

Inoue, S., and Kitajima, K. (2006). KDN (deaminated neuraminic acid): Dreamful past and exciting future of the newest member of the sialic acid family. *Glycoconjugate Journal, 23*, 277-290.

Jacobs, T., Erdmann, H., and Fleischer, B. (2010). Molecular interaction of Siglecs (sialic acid-binding Ig-like lectins) with sialylated ligands on Trypanosoma cruzi. *European Journal of Cell Biology, 89*, 113-116.

Janson, G., Zhang, C., Prado, M. G., and Paiardini, A. (2016). PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics*.

Jeffrey, G., and Jeffrey, G.. (1997). *An introduction to hydrogen bonding* (Vol. 32): Oxford university press New York.

Jones, F., Stanley, A., Oliveira, A., Rohl, A. L., Reyhani, M. M., Parkinson, G. M., and Ogden, M. I. (2003). The role of phosphonate speciation on the inhibition of barium sulfate precipitation. *Journal of Crystal Growth, 249*, 584-593.

Juhl, M., Laursen, S. L. R., Huang, Y. X., Nielsen, D. U., Daasbjerg, K., and Skrydstrup, T. (2017). Correction to Copper-Catalyzed Carboxylation of Hydroborated Disubstituted Alkenes and Terminal Alkynes with Cesium Fluoride (vol 7, pg 1392, 2017). *Acs Catalysis, 7*, 2247-2247. doi:

Kabachnik, M., and Medved, T.. (1954). New method for the synthesis of 1-aminoalkyl-phosphonic and phosphinic acids Communication 4. Synthesis of (1-aminoalkyl) phenylphosphinic acids. *Russian Chemical Bulletin, 3*(6), 893-900.

Kashif, M., Moreno-Herrera, A., Lara-Ramirez, E. E., Ramirez-Moreno, E., Bocanegra-Garcia, V., Ashfaq, M., and Rivera, G. (2017). Recent developments in trans-sialidase inhibitors of Trypanosoma cruzi. *J Drug Target*, 1-51.

Kendrew, J. C, Bodo, G, D., Howard, P. R., Wyckoff, H., and Phillips, D. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature, 181*, 662-666.

Khorlin, A. Y., Privalova, I. M., Zakstelskaya, L. Y., Molibog, E. V., and Evstigne.N. (1970). Synthetic Inhibitors of Vibrio-Cholerae Neuraminidase and Neuraminidases of Some Influenza Virus Strains. *Febs Letters, 8*, 17.

Kim, J. H., Ryu, H. W., Shim, J. H., Park, K. H., and Withers, S. G. (2009). Development of new and selective Trypanosoma cruzi trans-sialidase inhibitors from sulfonamide chalcones and their derivatives. *Chembiochem, 10*, 2475-2479.

Kiss, T., Lazar, I., and Kafarski, P. (1994). Chelating tendencies of bioactive aminophosphonates. *Met Based Drugs, 1*, 247-264.

Lara-Ramirez, E. E., Lopez-Cedillo, J. C., Nogueda-Torres, B., Kashif, M., Garcia-Perez, C., Bocanegra-Garcia, V., Rivera, G. (2017). An in vitro and in vivo evaluation of new potential trans-sialidase inhibitors of Trypanosoma cruzi predicted by a computational drug repositioning method. *Eur J Med Chem, 132*, 249-261.

Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A., and Murshudov, G. N. (2012). JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr, 68*, 431-440.

Lee, B. Y., Bacon, K. M., Bottazzi, M. E., and Hotez, P. J. (2013). Global economic burden of Chagas disease: a computational simulation model. *Lancet Infect Dis, 13*, 342-348.

Lee, B. Y., Bacon, K. M., Connor, D. L., Willig, A. M., and Bailey, R. R. (2010). The potential economic value of a Trypanosoma cruzi (Chagas disease) vaccine in Latin America. *PLoS Negl Trop Dis, 4*, e916.

Li, S. C., Squires, C. L., and Squires, C. (1984). Antitermination of E. coli rRNA transcription is caused by a control region segment containing lambda nut-like sequences. *Cell, 38*, 851-860.

Liu, H., and Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol, 8*, 91.

Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008). BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr, 64*, 125-132.

Lu, Y., Liu, Y., Xu, Z., Li, H., Liu, H., and Zhu, W. (2012). Halogen bonding for rational drug design and new drug discovery. *Expert Opin Drug Discov, 7*, 375-383.

Luque, I., and Freire, E. (2000). Structural stability of binding sites: consequences for binding affinity and allosteric effects. *Proteins, Suppl 4*, 63-71.

Machado, F. S., Jelicks, L. A., Kirchhoff, L. V., Shirani, J., Nagajyothi, F., Mukherjee, S., Tanowitz, H. B. (2012). Chagas heart disease: report on recent developments. *Cardiol Rev, 20*, 53-65.

Maguire, J. H. (2006). Chagas' disease--can we stop the deaths? *N Engl J Med, 355*, 760-761.

Marathe, B. M., Leveque, V., Klumpp, K., Webster, R. G., and Govorkova, E. A. (2013). Determination of neuraminidase kinetic constants using whole influenza virus preparations and correction for spectroscopic interference by a fluorogenic substrate. *PLoS One, 8*, e71401.

Marshall, M. S., Steele, R. P., Thanthiriwatte, K. S., and Sherrill, C. D. (2009). Potential energy curves for cation-pi interactions: off-axis configurations are also attractive. *J Phys Chem A, 113*, 13628-13632.

Martinez, C. R., and Iverson, B. L. (2012). Rethinking the term "pi-stacking". *Chemical Science, 3*, 2191-2201.

Mauro, A., Murali, N. (2010). W. I. P. Organization.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr, 40*, 658-674.

McLaughlin, M. M., Skoglund, E. W., and Ison, M. G. (2015). Peramivir: an intravenous neuraminidase inhibitor. *Expert Opinion on Pharmacotherapy, 16*, 1889-1900.

McPherson, A., and Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun, 70*, 2-20.

Mitchell, F. L., Miles, S. M., Neres, J., Bichenkova, E. V., and Bryce, R. A. (2010). Tryptophan as a molecular shovel in the glycosyl transfer activity of Trypanosoma cruzi trans-

sialidase. *Biophysical journal, 98*(9), L38-L40.

Monteiro, V. G., Soares, C. P., and de Souza, W. (1998). Host cell surface sialic acid residues are involved on the process of penetration of Toxoplasma gondii into mammalian cells. *Fems Microbiology Letters, 164*(2), 323-327.

Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr, 65*, 1074-1080.

Morillo, C. A., Marin-Neto, J. A., Avezum, A., Sosa-Estani, S., Rassi, A., Rosas, F., Investigators, BENEFIT. (2015). Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy. *New England Journal of Medicine, 373*, 1295-1306.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem, 30*(16), 2785-2791.

Morrot, A. (2013). The Role of Sialic Acid-Binding Receptors (Siglecs) in the Immunomodulatory Effects of Trypanosoma cruzi Sialoglycoproteins on the Protective Immunity of the Host. *Scientifica (Cairo), 2013*, 965856.

Moustafa, I., Connaris, H., Taylor, M., Zaitsev, V., Wilson, J. C., Kiefel, M. J., Taylor, G. (2004). Sialic acid recognition by Vibrio cholerae neuraminidase. *Journal of Biological Chemistry, 279*(39), 40819-40826.

Mu, X. J., Lei, M. Y., Zou, L. P., and Zhang, W. (2006). Microwave-assisted solvent-free and catalyst-free Kabachnik-Fields reactions for alpha-amino phosphonates. *Tetrahedron Letters, 47*(7), 1125-1127.

Munoz, C., Zulantay, I., Apt, W., Ortiz, S., Schijman, A. G., Bisio, M., Solari, A. (2013). Evaluation of Nifurtimox Treatment of Chronic Chagas Disease by Means of Several Parasitological Methods. *Antimicrobial Agents and Chemotherapy, 57*(9), 4518-4523.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 355-367.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr, 53*(Pt 3), 240-255.

Nóbrega, A. (2009). Oral Transmission of Chagas Disease by Consumption of Açaí Palm Fruit, Brazil-Volume 15, Number 4-April 2009-Emerging Infectious Disease journal-CDC.

Nardy, A. F., Freire-de-Lima, C. G., Perez, A. R., and Morrot, A. (2016). Role of Trypanosoma cruzi Trans-sialidase on the Escape from Host Immune Surveillance. *Front Microbiol, 7*, 348.

Naydenova, E. D., Todorov, P. T., and Troev, K. D. (2010). Recent synthesis of aminophosphonic acids as potential biological importance. *Amino Acids, 38*(1), 23-30.

Neres, J., Bonnet, P., Edwards, P. N., Kotian, P. L., Buschiazzo, A., Alzari, P. M., Douglas, K. T. (2007). Benzoic acid and pyridine derivatives as inhibitors of Trypanosoma cruzi trans-sialidase. *Bioorganic and Medicinal Chemistry, 15*, 2106-2119.

Neres, J., Brewer, M. L., Ratier, L., Botti, H., Buschiazzo, A., Edwards, P. N., Douglas, K. T. (2009). Discovery of novel inhibitors of Trypanosoma cruzi trans-sialidase from in silico screening. *Bioorg Med Chem Lett, 19*(3), 589-596.

Noireau, F., Diosque, P., and Jansen, A. M. (2009). Trypanosoma cruzi: adaptation to its vectors and its hosts. *Vet Res, 40*, 26.

Oliveira, I. A., Goncalves, A. S., Neves, J. L., von Itzstein, M., and Todeschini, A. R. (2014). Evidence of ternary complex formation in Trypanosoma cruzi trans-sialidase catalysis. *J Biol Chem, 289*, 423-436.

Otwinowski, Z., Minor, W., and *et al*. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol, 276*, 307-326.

Palese, P., Bucher, D., and Kilbourne, E. D. (1973). Applications of a synthetic neuraminidase substrate. *Appl Microbiol, 25*(2), 195-201.

Paris, G., Cremona, M. L., Amaya, M. F., Buschiazzo, A., Giambiagi, S., Frasch, A. C., and Alzari, P. M. (2001). Probing molecular function of trypanosomal sialidases: single point mutations can change substrate specificity and increase hydrolytic activity. *Glycobiology, 11*(4), 305-311.

Paris, G., Ratier, L., Amaya, M. F., Nguyen, T., Alzari, P. M., and Frasch, A. C. (2005). A sialidase mutant displaying trans-sialidase activity. *J Mol Biol, 345*, 923-934.

Pereira, P. C., and Navarro, E. C. (2013). Challenges and perspectives of Chagas disease: a review. *J Venom Anim Toxins Incl Trop Dis, 19*, 34.

Perez, C. J., Lymbery, A. J., and Thompson, R. A. (2015). Reactivation of Chagas disease:

implications for global health. *Trends in parasitology, 31*(11), 595-603.

Pinazo, M. J., Munoz, J., Posada, E., Lopez-Chejade, P., Gallego, M., Ayala, E., Gascon, J. (2010). Tolerance of benznidazole in treatment of Chagas' disease in adults. *Antimicrob Agents Chemother, 54*, 4896-4899.

Podlipaev, S. (2001). The more insect trypanosomatids under study-the more diverse Trypanosomatidae appears. *Int J Parasitol, 31*(5-6), 648-652.

Potier, M., Mameli, L., Belisle, M., Dallaire, L., and Melancon, S. B. (1979). Fluorometric assay of neuraminidase with a sodium (4-methylumbelliferyl-alpha-D-N-acetylneuraminate) substrate. *Anal Biochem, 94*, 287-296.

Prinz, H. (2010). Hill coefficients, dose-response curves and allosteric mechanisms. *J Chem Biol, 3*, 37-44.

Pritt, B. (2011). Trypanosoma brucei trypomastigotes, avaliable from: http:// parasitewonders.blogspot.co.uk / 2011/01/answer-to-case-147.html

Rassi, A., Jr., Rassi, A., and Marin-Neto, J. A. (2010). Chagas disease. *Lancet, 375*, 1388-1402.

Resende, R. (2010). Synthesis of novel sialidase inhibitors to target influenza A virus and Chagas' disease. University of Bath.

Riley, R. J., Grime, K., and Weaver, R. (2007). Time-dependent CYP inhibition. *Expert Opin Drug Metab Toxicol, 3*(1), 51-66.

Rosenberg, C. S., Zhang, W., Bustamante, J. M., and Tarleton, R. L. (2016). Long-Term Immunity to Trypanosoma cruzi in the Absence of Immunodominant trans-Sialidase-Specific CD8+ T Cells. *Infect Immun, 84*(9), 2627-2638.

Rupp, B. (2015). Reviewing biomolecular crystallography proposals: time for a paradigm change. *Trends Biochem Sci, 40*(8), 419-421.

Rupp, B. (2009). *Biomolecular crystallography: principles, practice, and application to structural biology*: Garland Science.

Ryabukhin, S. V., Panov, D. M., Shivanyuk, A. N., Plaskon, A. S., Zarudnitskiy, E. N., and Lukin, O. (2014). A One-Pot, Three-Step Synthesis of a-Aminophosphonic Acids. *Synthesis-Stuttgart, 46*, 2079-2084.

Salamanca-Dejour, D., Blanchet, D., Aznar, C., La Ruche, G., Jeannel, D., and Gastellu-Etchegorry, M. (2012). [Chagas disease (American trypanosomiasis) in France]. *Med*

Mal Infect, 42, 344-348.

Samraj, A. N., Laubli, H., Varki, N., and Varki, A. (2014). Involvement of a non-human sialic Acid in human cancer. *Front Oncol, 4*, 33.

San Francisco, J., Barría, I., Gutiérrez, B., Neira, I., Muñoz, C., Sagua, H., Catalán, A. (2017). Decreased cruzipain and gp85/trans-sialidase family protein expression contributes to loss of Trypanosoma cruzi trypomastigote virulence. *Microbes and Infection, 19*(1), 55-61.

Sathe, A. A., Hartline, D. R., and Radosevich, A. T. (2013). A synthesis of alpha-amino acids via direct reductive carboxylation of imines with carbon dioxide. *Chemical Communications, 49*(44), 5040-5042.

Scapin, G. (2013). Molecular replacement then and now. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 11), 2266-2275.

Schenkman, S., Decarvalho, L. P., and Nussenzweig, V. (1992). Trypanosoma-Cruzi Trans-Sialidase and Neuraminidase Activities Can Be Mediated by the Same Enzymes. *Journal of Experimental Medicine, 175*, 567-575.

Schenkman, S., Jiang, M. S., Hart, G. W., and Nussenzweig, V. (1991). A novel cell surface trans-sialidase of Trypanosoma cruzi generates a stage-specific epitope required for invasion of mammalian cells. *Cell, 65*, 1117-1125.

Schmitt, B. H., Rosenblatt, J. E., and Pritt, B. S. (2012). Laboratory diagnosis of tropical infections. *Infectious disease clinics of North America, 26*(2), 513-554.

Schwender, C. F., Beers, S. A., Malloy, E. A., Cinicola, J. J., Wustrow, D. J., Demarest, K. D., and Jordan, J. (1996). Phenylphosphonic acid inhibitors of human prostatic acid phosphatase. *Bioorganic and Medicinal Chemistry Letters, 6*, 311-314.

Severi, E., Hood, D. W., and Thomas, G. H. (2007). Sialic acid utilization by bacterial pathogens. *Microbiology, 153*, 2817-2822.

Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallogr A, 64*, 112-122.

Silva, B. L., Filho, J. D. S., Andrade, P., Carvalho, I., and Alves, R. J. (2014). Design, synthesis and enzymatic evaluation of 3-O-substituted aryl beta-D-galactopyranosides as inhibitors of Trypanosoma cruzi trans-sialidase. *Bioorganic and Medicinal Chemistry Letters, 24*, 4529-4532.

Smith, B. J., Colman, P. M., Von Itzstein, M., Danylec, B., and Varghese, J. N. (2001). Analysis

of inhibitor binding in influenza virus neuraminidase. *Protein Sci, 10*, 689-696.

Sodre, C. L., Chapeaurouge, A. D., Kalume, D. E., de Mendonca Lima, L., Perales, J., and Fernandes, O. (2009). Proteomic map of Trypanosoma cruzi CL Brener: the reference strain of the genome project. *Arch Microbiol, 191*, 177-184.

Stencel-Baerenwald, J. E., Reiss, K., Reiter, D. M., Stehle, T., and Dermody, T. S. (2014). The sweet spot: defining virus-sialic acid interactions. *Nature Reviews Microbiology, 12*, 739-749.

Steverding, D. (2014). The history of Chagas disease. *Parasit Vectors, 7*, 317.

Stanaway, J. D. and Gregory R. "The burden of Chagas disease: estimates and challenges." Global heart 10.3 (2015): 139-144.

Stoco, P. H., Wagner, G., Talavera-Lopez, C., Gerber, A., Zaha, A., Thompson, C. E., Grisard, E. C. (2014). Genome of the avirulent human-infective trypanosome--Trypanosoma rangeli. *PLoS Negl Trop Dis, 8*, e3176.

Streicher, H., and Busse, H. (2006). Building a successful structural motif into sialylmimetics—cyclohexenephosphonate monoesters as pseudo-sialosides with promising inhibitory properties. *Bioorganic and medicinal chemistry, 14*(4), 1047-1057.

Su, H., Altucci, L., and You, Q. (2008). Competitive or noncompetitive, that's the question: research toward histone deacetylase inhibitors. *Mol Cancer Ther, 7* 1007-1012.

Sutar, R. L., and Joshi, N. N. (2013). Systematic evaluation of a few proline derivatives as catalysts for a direct aldol reaction. *Tetrahedron-Asymmetry, 24*, 43-49.

Tajbakhsh, M., Heydari, A., Alinezhad, H., Ghanei, M., and Khaksar, S. (2008). Coupling of aldehydes, amines, and trimethyl phosphite promoted by amberlyst-15: Highly efficient synthesis of alpha-aminophosphonates. *Synthesis-Stuttgart*, 352-354.

Tan, T. T., and Coussens, L. M. (2007). Humoral immunity, inflammation and cancer. *Curr Opin Immunol, 19*, 209-216.

Tartof, K., and Hobbs, C. (1987). Improved media for growing plasmid and cosmid clones. *Focus, 9*(2), 12.

Telford, J. C. (2014). Structural studies on the sialidase from Aspergillus fumigatus, and, Fragment based drug discovery against the trans-sialidase from Trypanosoma cruzi. (PhD), University of St Andrews.

Todeschini, A. R., Nunes, M. P., Pires, R. S., Lopes, M. F., Previato, J. O., Mendonca-Previato, L., and DosReis, G. A. (2002). Costimulation of host T lymphocytes by a trypanosomal trans-sialidase: Involvement of CD43 signaling. *Journal of Immunology, 168*(10), 5192-5198.

Torrecilhas, A. C. T., Tonelli, R. R., Pavanelli, W. R., da Silva, J. S., Schumacher, R. I., de Souza, W., Alves, M. J. M. (2009). Trypanosoma cruzi: parasite shed vesicles increase heart parasitism and generate an intense inflammatory response. *Microbes and Infection, 11*, 29-39.

Tribulatti, M. V., Mucci, J., Van Rooijen, N., Leguizamon, M. S., and Campetella, O. (2005). The trans-sialidase from Trypanosoma cruzi induces thrombocytopenia during acute Chagas' disease by reducing the platelet sialic acid contents. *Infection and Immunity, 73*, 201-207.

Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem, 31*, 455-461.

Turnbull, W. B., and Daranas, A. H. (2003). On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc, 125*, 14859-14866.

Tyler, K. M., and Engman, D. M. (2001). The life cycle of Trypanosoma cruzi revisited. *Int J Parasitol, 31*(5-6), 472-481.

van der Rest, M. E., Lange, C., and Molenaar, D. (1999). A heat shock following electroporation induces highly efficient transformation of Corynebacterium glutamicum with xenogeneic plasmid DNA. *Appl Microbiol Biotechnol, 52*(4), 541-545.

Vandekerckhove, F., Schenkman, S., Decarvalho, L. P., Tomlinson, S., Kiso, M., Yoshida, M., Nussenzweig, V. (1992). Substrate-Specificity of the Trypanosoma-Cruzi Trans-Sialidase. *Glycobiology, 2*, 541-548.

Varki, A. (2010). Colloquium paper: uniquely human evolution of sialic acid genetics and biology. *Proc Natl Acad Sci U S A, 107 Suppl 2*, 8939-8946.

Varki, A., and Gagneux, P. (2012). Multifarious roles of sialic acids in immunity. *Glycobiology of the Immune Response, 1253*, 16-36.

Varki, A. (2009). *Essentials of glycobiology* (pp. xxix, 784 p.). Retrieved from http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=glyco2

Varki, N. M., and Varki, A. (2007). Diversity in cell surface sialic acid presentations: implications for biology and disease. *Lab Invest, 87*(9), 851-857.

Vogel, A. I., and Furniss, B. S. (1989). *Vogel's textbook of practical organic chemistry* (5th ed / ed.). Harlow: Prentice Hall.

Waespy, M., Gbem, T. T., Elenschneider, L., Jeck, A. P., Day, C. J., Hartley-Tassell, L., Kelm, S. (2015). Carbohydrate Recognition Specificity of Trans-sialidase Lectin Domain from Trypanosoma congolense. *Plos Neglected Tropical Diseases, 9*(10).

Wang, P. C., Fang, J. M., Tsai, K. C., Wang, S. Y., Huang, W. I., Tseng, Y. C., Wong, C. H. (2016). Peramivir Phosphonate Derivatives as Influenza Neuraminidase Inhibitors. *Journal of Medicinal Chemistry, 59*, 5297-5310.

Watts, A. G., Damager, I., Amaya, M. L., Buschiazzo, A., Alzari, P., Frasch, A. C., and Withers, S. G. (2003). Trypanosoma cruzi trans-sialidase operates through a covalent sialyl-enzyme intermediate: Tyrosine is the catalytic nucleophile. *Journal of the American Chemical Society, 125*, 7532-7533.

Watts, A. G., Oppezzo, P., Withers, S. G., Alzari, P. M., and Buschiazzo, A. (2006). Structural and kinetic analysis of two covalent sialosyl-enzyme intermediates on Trypanosoma rangeli sialidase. *J Biol Chem, 281*, 4149-4155.

Weiner, M. P., Costa, G. L., Schoettlin, W., Cline, J., Mathur, E., and Bauer, J. C. (1994). Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene, 151*(1-2), 119-123.

Wells, E. A. (1972). The importance of mechanical transmission in the epidemiology of nagana: a review. *Trop Anim Health Prod, 4*(2), 74-88.

White, Clinton L, Janakiraman, Musiri N, Laver, Graeme W, Philippon, Cédric, Vasella, Andrea, Air, Gillian M, and Luo, Ming. (1995). A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. *Journal of molecular biology, 245*(5), 623-634.

WHO. (2002). Control of Chagas disease. *World Health Organ Tech Rep Ser, 905*, i-vi, 1-109, back cover.

WHO. (2017). WHO Trypanosomiasis, human African (sleeping sickness). from http://www.who.int/mediacentre/factsheets/fs259/en/

Wiederschain, G. Y. (2009). Essentials of glycobiology. *Biochemistry (Moscow), 74*(9), 1056-

1056.

Wilbrink, M. H., Geert, L., Sander S, Sanders, Sallomons, P. Erik, Hage, Johannes A, Kamerling, Johannis P. (2014). Galactosyl-lactose sialylation using Trypanosoma cruzi trans-sialidase as the biocatalyst and bovine κ-casein-derived glycomacropeptide as the donor substrate. *Applied and environmental microbiology, 80*, 5984-5991.

Winter, G., Lobley, C. M., and Prince, S. M. (2013). Decision making in xia2. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 7), 1260-1273.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *Febs Journal, 275*, 1-21.

Yang, L., Connaris, H., Potter, J. A., and Taylor, G. L. (2015). Structural characterization of the carbohydrate-binding module of NanA sialidase, a pneumococcal virulence factor. *BMC Struct Biol, 15*, 15.

Yoshida, N., Dorta, M. L., Ferreira, A. T., Oshiro, M. E. M., Mortara, R. A., AcostaSerrano, A., and Favoreto, S. (1997). Removal of sialic acid from mucin-like surface molecules of Trypanosoma cruzi metacyclic trypomastigotes enhances parasite-host cell interaction. *Molecular and Biochemical Parasitology, 84*, 57-67.